This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

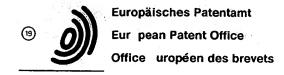
Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problems Mailbox.





1) Publication number:

0 518 313 A2

(12)

Market Market Street

EUROPEAN PATENT APPLICATION

(21) Application number: 92109812.5

② Date of filing: 11.06.92

(5) Int. Cl.5: **C12N 15/51**, C07K 15/00, C12Q 1/70, A61K 39/29, G01N 33/576

A request for correction of the sequence listing has been filed pursuant to Rule 88 EPC. A decision on the request will be taken during the proceedings before the Examining Division (Guidelines for Examination in the EPO, A-V, 2.2).

- Priority: 11.06.91 JP 139268/91
 12.07.91 JP 172794/91
 07.10.91 JP 287008/91
 16.12.91 JP 332329/91
 20.04.92 JP 99957/92
- 43 Date of publication of application: 16.12.92 Bulletin 92/51
- Designated Contracting States:
 AT BE CH DE DK ES FR GB GR IT LI LU MC
 NL PT SE

Applicant: MITSUBISHI KASEI CORPORATION 5-2, Marunouchi 2-chome Chiyoda-ku Tokyo 100(JP)

10-406, Machida-cooptown, 2-11-3, Ogawa Machida-shi, Tokyo-to(JP)
Inventor: Honda, Yoshikazu
2-3-16, Nishimikado
Kamakura-shi, Kanagawa-ken(JP)
Inventor: Takahashi, Kazuhiro
10-201, Popuragaoka-coop, 2-10-1, Naruse Machida-shi, Tokyo-to(JP)
Inventor: Murakami, Tomoko
3-2-9, Narusegaoka
Machida-shi, Tokyo-to(JP)
Inventor: Teranishi, Yutaka
1-14-2, Takane, Sagamihara-shi

Kanagawa-ken(JP) Inventor: Hayashi, Norio 1-1-112, Suimeidai

Kawanishi-shi, Hyogo-ken(JP)

- Representative: Hansen, Bernd, Dr. Dipl.-Chem. et al Hoffmann, Eitle & Partner Patent- und Rechtsanwälte Arabeliastrasse 4 Postfach 81 04 20 W-8000 München 81(DE)
- Gene of hepatitis C virus or fragment thereof, polypeptide encoded by the same.
- A novel gene encoding HCV polypeptide including HCV-associated antigen, a polypeptide encoded by the same, an expression vector containing the gene, a transformant transformed with the vector, a process for producing HCV polypeptid by culturing the transformant, which polypeptide produced by the process is useful for s rodiagnosis of hepatitis C and for the preparation of vaccine against hepatitis C virus.

Field of the invention

5

This invention relates to an isolated gene encoding a polypeptide of human hepatitis C virus (hereinafter, referred to as HCV), or a fragment thereof, and a polypeptide encoded thereby.

Background of the invention

Hepatitis viruses A, B and D have been identified and the serodiagnosis for each virus has been established before the present invention. However, there was at least one hepatitis whose cause remained unknown (Digestive Diseases and Sciences, 31: 122S-132S (1986); and Seminars in Liver Diseases, 6: 56-66 (1986)).

Serodiagnosis for hepatitis A virus (HAV) or hepatitis B virus (HBV) has been established and clinically employed since middle of 1970's, which revealed that most of the blood-transfusion-associated hepatitis are caused by unknown pathogen(s) other than viruses capable of growing in hepatocytes, such as HAV or HBV. The hepatitis caused by unknown pathogen was designated as "non-A, non-B hepatitis (NANBH)". In the United States, the incidence of hepatitis following the "transfusion" is about 1 to 10% of the total patients undergone transfusion, and more than 90% of said post-transfusional hepatitis are reported to be NANBH (Jikken Igaku, 8,3: 15-18 (1990)). In Japan, about 200,000 patients, corresponding to about 10 - 20% of those undergone transfusion, are suffering from the post-transfusional hepatitis every year, and about 95% of them are diagnosed as NANBH. Furthermore, about 300,000 people are diagnosed as sporadic hepatitis every year and about 40 to 50% of them are considered to be NANBH. There are also epidemic NANBH in Japan. Although infectious route for NANBH has not been established in contrast with hepatitis A or B, it is likely different from those for hepatitis A and B (Jikken Igaku, 8, 3: 13-14 (1990)).

Chiron Corp. (May, 1988) has succeeded in isolating a gene fragment of a virus responsible for NANBH by means of an unique technique quite different from conventional ones and designated said virus as hepatitis C virus (HCV). Many researchers followed the work and sequenced the entire gene encoding both of non-structural and structural proteins of HCV (Shimotohno et al., Proc. Natl. Acad. Sci. USA, 87: 9524-9526 (1990); and Takamizawa et al., Journal of Virology 65, 3: 1105-1113 (1991)).

Many Patent Applications directed to HCV gene have been done so far, for example, European Patent Publication Nos.318216, 388232, 398748, 419182, 450931, 464287, 463848, 468657, WO 91/01376, WO 91/15516 and British Patent No.2239245, and the like.

Chiron corp. and Ortho, Inc. have developed an Enzyme-linked Immunosorbent Assay (ELISA) for HCV and a kit therefor, using a recombinant antigen (clone C100-3) which was obtained by transforming yeast cells with an expression plasmid encoding a fused peptide comprising a human super-oxide dismutase and a 363 amino acid polypeptide encoded by a gene encoding a region from NS3 to NS4 encoding a part of non-structural protein, growing transformants under a condition to allow the transformants express said fused peptide (WO No. 89/04669; and European Patent Publication No.318216).

The Japanese Welfare Ministry (KOSEI-SHO), leading other nations in the world, decided to introduce said Chiron's kit into the screening and detection of anti-HCV antibody and the import thereof started on December 26, 1989. From the next day, the Japanese Red Cross began screening for anti-HCV antibody in blood offered by volunteers using the kit. About 1.7 million of people are estimated to undergo blood transfusion yearly. Before the screening, the incidence of post-transfusional hepatitis among them was about 12.3% (about 173,000), and thereafter, it reduced to about 3%.

As an outstanding and critical feature, the variability of HC-associated antigen is often suggested. For instance, homology in amino acid and base sequences between C100-3 clone and a clone obtained in Japan was reported to be about 80%. The difference is only 20% though, it can affect on the accuracy of the detection of HCV. In another aspects, homology between HC-associated antigens varies from a region to region, for example, it is only 70% regarding the all or a part of NS1, NS2, NS3 and NS5 regions (according to the designation by Chiron Corp.), which indicates that some substances may be overlooked by Chiron's kit. As is often the case with virus, especially that has RNA genome, a genetic mutation occurs at a high frequency, which leads to a change in antigen determinant sites. As a result, HC-associated antigen presented by antigen-presenting cells in serum and antibody raised against it also change in the course of disease.

The another kit provided by Ortho, Inc. is accurate in detecting anti-HCV antibodies raised during a restricted period of diseas, that is, antibodies raised during a period while the disease progresses from an acute stage to a chronic stage, which begins about 24.7 weeks after the infection (SAISHIN IGAKU, 45, 12: 2331-2336 (19909); IGAKUNOAYUMI, 151: 892-896). Thus, the Ortho's kit is not effective for detecting antibodies raised against all the HC-associated antigens throughout the disease, especially those presented

during acute and chronic stag s.

Accordingly, an assay method useful for the detection of any anti-HCV antibodies raised against various HC-associated antigens exist in serum of a patient throughout the disease has been needed. HCV is detectable in hepatocytes of patients in various phases of disease, including acute, chronic hepatitis, hepatocirrhosis, and hepatoma. Recently, interests are concentrated on the pathogenetic relationship between HCV infection and hepatoma because about 50 - 60% of patients of hepatoma are HCV positive. Although the pathogenetic relationship between HCV infection and hepatoma has not been established, it is generally accepted that there are some relationships between chronic hepatitis, hepatocirrhosis and hepatoma. Therefore, screening for anti-HCV antibody in serum of a subject susceptible to them may helpful for preventing such serious diseases. Thus, more accurate and efficient screening method, as well as serodiagnosis, is strongly desired to prevent HCV-related diseases. For this purpose, a reagent and a kit having a extended utility in, for example, the assay of serum of a variety of subjects including carriers of HCV without manifesting symptoms, patients suffering from HC of various stages, such as acute, chronic, or progressed hepatitis, is necessary. As the number of HCV-infected patients increases, the% of HCVcontamination in blood offered by volunteers increases. This causes a serious problem all over the world, for instance, the% of HCV positive blood in total blood offered by volunteers is about 10, 0.8, 1.5, and 1.2% in Japan, USA, Italy, and Spain, respectively (TANPAKUSITU, KAKUSAN, KOSO, 36, 10: 1679-1691 (1991)-). However, there are no effective methods for treating HCV infection, and therefore a method for detection of HCV in serum of suspected subjects is strongly required to prevent HCV-related diseases.

20 Summary of the Invention

As previously mentioned, HCV gene is extremely liable to vary and subtypes of HCV should differ to a great extent at various sites including surface antigenic sites and others responsible for the determination of significant features of HCV protein. As these mutated viruses induce hepatitis C of different symptoms depending on the type when infected to human, variants low in homology are considered to be different from each other.

In this regard, the present inventors isolated plural viruses which differ from each other in terms of amino acid and DNA sequences from sera of patients of HC (HC patients).

The present invention was established by isolating a novel hepatitis C virus, separating RNA encoding viral protein, converting RNA into cDNA using reverse transcriptase, and cloning and sequencing the resultant DNA. When the isolated DNA was transformed into host cells after ligating to an appropriate expression vector, transformants expressed HC-associated antigen.

DNA obtained by transcribing the RNA of HCV encodes recombinant antigen which is immunochemically the same as HCV-associated antigen. Therefore, for the purpose of the invention, the terms "cDNA", "DNA" and "gene" are used interchangeably, as far as they encode the same protein(s) or antigens as those encoded by RNA gene of HCV. As one of skill will easily appreciate, a DNA fragment encoding an epitopic site of HCV-associated antigen is also useful to produce a polypeptide capable of specifically reacting with anti-HCV antibody in the same manner as intact HC-associated antigen. Therefore such a DNA fragment is also useful for the purpose of the invention.

Thus, the present invention provides an isolated gene of a novel hepatitis C virus and a fragment thereof. The HCV gene and its fragment of the invention are useful for the development of a diagnostic method which is more accurate and effective than conventional ones in the detection of antibodies raised against a wide range of HCVs which have been hardly detected before the present invention. The gene and fragments thereof are also useful for the preparation of a novel vaccine.

In another aspects of the invention, an <u>in vitro</u> screening system for a substance capable of specifically suppressing or controlling a proteolytic processing of a precursor polypeptide of HCV can be obtained. The screening system can be established by analyzing viral protease intimately. The analysis can be carried out by synthesizing a + strand of RNA from a double-stranded DNA containing HCV-originated protease gene and its adjoining regions, producing a polypeptide comprising viral protease <u>in vitro</u>, characterizing said protease as to the activity, specificity, function, and the like.

In another aspects, the present invention provid s an in vivo screening system for the substance capable of suppressing the processing of viral precursor protein. The screening can be carried out using a transformant, for example, eucaryotic cells such as animal cells which have been transformed with DNA fragment of the invention, and can express a precursor polyp ptide of HCV and proc ss the product intracellularly.

Specifically, the present invention provides an isolated DNA (gene) encoding all or a part of polypeptide having an amino acid sequence of any of SEQ ID NO 1 to 43, 64 to 75 and 101 to 104, or fragment thereof.

The present invention further provides polypeptide having all or a part of amino acid sequence of any of SEQ ID. Nos. 1 to 43, 64 to 75 and 101 to 104.

The polypeptides of the invention have an ability to immunochemically and specifically react with antiserum obtained from patients suffering from h patitis C.

As the amino acid and DNA sequences of polypeptide of HCV are determined, it is easy to obtain active derivatives of viral protein which falls within the scope of the present invention by conventional methods which leads to the insertion, deletion, replacement or addition of amino acids without changing the specific reactivity with sera from patients suffering from hepatitis C. This can be conducted by, for example, a site-specific mutagenesis of DNA.

Therefore, the present invention also provides active derivatives of HCV protein obtained by conventional methods, and DNA fragments encoding it, which can immunochemically react with antiserum raised against HC-associated antigen.

In this regard, the present invention provides a polypeptide fragment having a modified amino acid sequence derived from polypeptides having amino acid sequence of SEQ ID NO 1 - 104, and being capable of reacting with serum of HC patients with a different specificity, for example, those claimed in Claims 119, 121, 123, 125, 127, 129, 131 - 136, 138, 140 - 154, 157 - 179, 185 - 199, wherein the modification has been done by deletion, insertion, modification or addition of amino acid(s) subject to that the ability to react with antiserum from HC patients is not decreased:

Furthermore, the present invention provides an expression vector which comprises DNA shown by either of SEQ ID NO 1 to 43, 64 to 75 and 101 to 104 or a fragment thereof and has an ability to allow a host cell to express said DNA when transformed into the same.

The present invention also provides a transformant transformed with the expression vector.

The present invention further provides a method for preparing HCV protein or HCV-associated antigen by culturing a transformant in a medium and recovering the product from the cultured broth.

Definition

25

35

40

45

For the purpose of the invention, the following terms are defined below.

HCAg: HCV- or HC-associated antigen. For the purpose of the invention, as hepatitis C is caused by HCV, the terms "HC-associated antigen" and "HCV-associated antigen" are used exchangeably.

HCAb: antibody raised against HCV-associated antigen.

HCV protein or HCV polypeptide: protein or polypeptide encoded by HCV gene.

HCV gene: generally, it is RNA gene of HCV. However, for the purpose of the invention, it refers to gene encoding HCV polypeptide or protein encoded by RNA gene. Therefore, the terms "gene", "cDNA", and "DNA" obtained from RNA gene are used exchangeably.

Recombinant HCAg: a product (protein or polypeptide, including glycosylated ones) produced in host cells transformed by DNA of the invention and is capable of immunochemically reacting with HCAb.

Recombinant polypeptide: polypeptide expressed by host cells transformed by HCV gene of the invention.

HC patient: a patient suffering from hepatitis C.

Detailed Description of the Invention

[1] Gene Encoding Core-envelope Region

(1) Preparation of cDNA clone of SEQ ID NO 1 - 12 and sequencing thereof

The cDNA clones of SEQ ID NO 1 - 12 which encode a novel polypeptide of core-envelope region of HCV protein were cloned from serum from HC patients as follows.

The cloning and sequencing of cDNA encoding HCV polypeptide can be carried out using any of known methods. However, it is hardly accomplished by known "Okayama-Berg" or "Gubler-Hoffman" method because the content of HCV in serum is only a slight amount and HCV gene is liable to variation. The present inventors succeeded in the cloning of gene from a slight amount of serum as will be hereinafter described in Example 1. Briefly, it was conducted by extracting nucleic acids from a serum of a patient suffering from HC. It is preferable to use serum showing OD value of 3.5 on a screening kit of Ortho. Before the extraction, it is desirable to add tRNA or polyribonucleoside to the serum as a carrier for viral RNA. For the purpose of the invention, tRNA is preferable because the degradation of RNA can be easily detected, at least after the addition of tRNA, by monitoring the existence of a sufficient amounts of tRNA having an

intact length on electrophoresis.

The r sultant RNA is converted into cDNA using transcriptase in the presence of an appropriate oligonucleotide primer. The cDNA is then cloned and amplified by modified polymerase chain reaction (PCR) (Saiki et al., Nature 324: 126 (1986)) in the presence of a pair of primers. Although commercially available random primers can be used in the PCR, synthetic primers having the following base sequences are suitable for the present invention.

Synthetic Primers

5' 3'

S1:CTCCACCATAGATCACTCC(SEQ ID NO:105)

S2:AGGTCTAGTAGACCGTGC(SEQ ID NO:106)

S3:AGGAAGACTTCCGAGCGG(SEQ ID NO:107)

S4:CGTGAACTATGCAACAGGG(SEQ ID NO:108)

AS1:ACCGCTCGGAAGTCTTCC(SEQ ID NO:109)

AS2:GGGCAAGTTCCCTGTTGC(SEQ ID NO:110)

AS3:GCTGGATTCTCTGAGACG(SEQ ID NO:111)

PCR can be conducted under appropriate conditions, for example, those described in Example 2 using the first complemental DNA (1st cDNA) as a template. The condition may vary depending on the primers used such as base sequence or combination, length to be amplified, or the like. Examples of pair of primers are: S1 - AS1; S1 - AS2; S1 - AS3; S2 - AS1; S2 - AS2; S2 - AS3; S3 - AS2; S3 - AS3; and S4 - AS3.

The minimum amount of serum required for the cloning described in Example 2 [2] varies depending on the content of virus in serum used, however, it may be about 5 to 7 μ I when the serum shows OD 3 or more on aforementioned Ortho's kit. The base sequence of cDNA obtained using random primer in the synthesis of the 1st strand cDNA was the same as that of cDNA obtained using antisence primer which was designed and synthesized (Example 8).

Thus, a region (clone N1-1) was obtained by two different methods. Three clones independently obtained from a serum of a patient using random primers are shown as a clone of SEQ ID NO 1. When synthetic DNA (S1 and AS1) was used as primers, two clones of three clones obtained independently have the same base sequence as that of SEQ ID NO 1 and one clone had a modified base sequence wherein three amino acids of SEQ ID NO 1 were changed, i.e., No.345 A to C, No.332 A to T, and No. 95 A to C, which shows that there are more than one virus in one patient.

The resultant DNA fragment is then subjected to the determination of base sequence. Generally, three clones obtained independently are employed and the base sequence of the both strands are determined to obtain an entire base sequence. The base sequence is conveniently determined using a fluorescence sequencer GENESIS 2000 (DUPONT) according to the protocol attached thereto. Alternatively, a conventional subcloning can be used when the DNA fragment consists more than 180 nucleotides or contains a region which is hardly determined by fluorescence sequencer.

Thus obtained base sequences are shown in SEQ ID NO 1 to 12.

For the purpose of the invention, a part of base sequences may be changed, for example, No. 345 A to C, No. 332 A to T, and No. 95 A to C, respectively.

(2) Expression of Polypeptides Encoded by Clones

50

55

DNA fragments of SEQ ID NO 1 to 12 can be used to produce a recombinant HCAg by constructing an expression vector containing DNA encoding a clone, by inserting the DNA into a known expression vector at an appropriate site of the vector, downstream from a promoter, using a well known method per se, and

introducing the expression vector harboring the DNA into a host cell such as Escherichia coli cell, yeast c ll, animal c ll or the like according to the method known to one of skill, culturing the transformant in a medium under an appropriate condition, and recovering a product from the cultured broth.

The present invention can be accomplished using any expression v ctors which have a promoter at an appropriate site to direct the expression of a DNA encoding HCV polypeptide or a fragment thereof. Expression vectors preferably contain promoter, ribosome binding (SD) sequence, gene encoding HCV polypeptide, transcription termination factor, and a regulator gene.

Expression vectors functional in microorganisms such as Escherichia coli, Bacillus subtilis or the like will preferably comprise promoter, ribosome binding (SD) sequence, HCV-associated-protein-encoding gene, transcription termination factor, and a regulator gene.

Examples of promoters include those derived from Escherichia coli or phages such as tryptophane synthetase (trp), lactose operon (lac), λ phage P_L and P_R , T_5 early gene P_{25} , P_{26} promoter and the like. These promoter may have modified or designed sequence for each expression vector such as pac promoter.

Although the SD sequence may be derived from Escherichia coli or phage, a sequence which has been designed to contain a consensus sequence consisting of more than 4 bases, which is complementary to the sequence at the 3' terminal region of 16S ribosome RNA, may also be used.

15

20

30

45

50

The transcription termination factor is not essential. However, it is preferable that an expression vector contains a ρ -independent factor such as lipoprotein terminator, trp operon terminator or the like.

Preferably, these sequences required for the expression of the a gene encoding HCAg originated from HCV are located, in an appropriate expression plasmid, in the order of promoter, SD sequence, said gene and transcription termination factor from 5' to 3' direction.

Typical example of expression vectors is commercially available pKK233-2 (Pharmacia). However, a series of plasmids pGEX (Pharmacia), which are provided for the expression of a fused protein, are also employable for the expression of HCAg-encoding gene of the present invention.

A suitable host cell such as Escherichia coli can be transformed with an expression vector comprising a DNA of the invention by any of known methods such as protocol provided by TOYOBO Japan (Example 3).

The cultivation of the transformants can be carried out using any of well known procedures in literatures such as Molecular Cloning, 1982, and the like. The cultivation is preferably conducted at a temperature from about 28°C to 42°C.

Expression vectors used for transforming other host cells, such as those derived from insects or animals including mammals, consist of substantially the same elements as those described in the above. However, there are certain preferable factors as follows.

When insect cells are used, a commercially available kit MAXBAC™ is employed according to the teaching of the supplier (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4). In this case, it is desirable to make a modification to reduce the distance between the promoter of polyhedrin gene and the initiation codon so as to improve the expression of the gene.

When animal cells are used as hosts, expression vectors preferably contain SV40 early promoter, SV40 late promoter, apolipoprotein E gene promoter, or the like. Specifically, known expression vectors such as pKCR (Proc. Natl. Acad. Sci. USA, 78: 1528 (1981)), pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), or the like may be employed after minimum modification, for example, an insertion of cloning site as described in a literature (Nature, 307: 604 (1984)), so that the resultant vectors maintain essential functions to serve as expression vectors.

Animal cells usable in the present invention are CHO cell, COS cell, mouse L cell, mouse C127 cell, mouse FM3A cell and the like.

The recombinant polypeptide expressed by host cells such as microorganisms including E. coli, insect cells and animal cells can be recovered from the cultured broth by known methods and identified by, for example, immunoreactions between the expressed product and antiserum obtained from HC patients using a conventional method such as Western blot analysis.

As the result, polypeptides having amino acid sequence of SEQ ID NO 1 to 12 were obtained as expression products of cDNA obtained from serum of HC patients and identified as HCAg. Among them, polypeptides having 191 amino acid sequence from No. 1 to No. 191 of SEQ ID NO 5, 6 and 8 are assumed to be polypeptides which were expressed and cleaved by proc ssing in insect cells. Thus, the sequences of SEQ ID NO 5, 6 and 8 comprise: from No. 174 to No. 188 (region A), amino acid sequence containing mainly hydrophobic amino acids having a large side chain of high r molecular weight; and at Nos. 189 and 191, alanine, a r sidue having a small side chain of lower molecular weight. This pattern of sequence keeps a feature of signal region which is recognized by signal peptidase in animal (including insect) cell. The 5'- and 3'- regions of said sequence contain many variations in amino acid sequence

resulting from variations in base sequence due to the replacement of a part of said sequence, when compared with known HCV genes cloned before the present invention. However, the r gions A and B contain less variations which indicates that the polypeptide may be cleaved at C-terminus of the No. 191 alanin by signal peptidase.

Polypeptide having amino acid sequence from No. 1 to No.191 of SEQ ID NO 5, 6, and 8 is assumed to be core or matrix protein on the basis of the homology between said amino acid sequence and a known sequence of core or matrix region of viral protein of Japanese encephalitis virus or yellow fever virus. The polypeptide comprising said 191 amino acid sequence is herein referred to as "core protein" or "core region".

Polypeptides having 18 amino acid sequence from No. 1 to No.18 of SEQ ID NO 1, 9, 10, 11 and 12, 34 amino acid sequence from No. 40 to No.73 of SEQ ID NO 3, and 35 amino acid sequence from No. 81 to No.115 of SEQ ID NO 3 are relatively highly hydrophilic and highly homologous to polypeptides having amino acid sequences deduced from known HCV genes cloned by Chiron, Shimotohno or Takamizawa (ibid) and are useful as HCV-associated antigenic peptide in diagnosis and/or for the preparation of vaccine.

Polypeptides having 18 amino acid sequence from No. 40 to No.57 of SEQ ID NO 4, and 12 amino acid sequence from No. 240 to No.251 of SEQ ID NO 4 are relatively highly hydrophilic and extremely low in homology with polypeptides having amino acid sequences deduced from known HCV genes cloned before the present invention and are useful as HCV-associated antigenic peptide in diagnosis. These polypeptides can be produced by chemical synthesis, as well as by DNA recombinant technique.

Furthermore, a polypeptide having 115 amino acid sequence from No. 1 to No. 115 of SEQ ID NO 3, corresponding to an epitopic region of core protein, and a polypeptide having 191 amino acid sequence from No. 1 to No.191 of SEQ ID NO 3, corresponding to the total region of core protein, can be produced in large scale by DNA recombinant technique and are useful as diagnostic reagent and/or vaccine.

Among them, a polypeptide having 192 amino acid sequence from No. 31 to No. 222 of SEQ ID NO 4 is assumed to be a polypeptide which was expressed and cleaved by processing in insect cells. Thus, the sequence of SEQ ID NO 4 comprises: from No. 13 to No. 29 (region A), amino acid sequence containing mainly hydrophobic amino acids having a side chain of higher molecular weight; at No. 30, alanine, a residue having a side chain of lower molecular weight; from No. 210 to No. 221 (region B), amino acid sequence containing mainly hydrophobic amino acids having a side chain of higher molecular weight; at No. 222, glycine, a residue having a side chain of lower molecular weight. This pattern of sequence keeps a feature of signal region which is recognized by signal peptidase in animal (including insect) cell.

The 5'- and 3'- regions of said sequence contain many variations in amino acid sequence resulting from those in base sequence due to the replacement of base sequences, when compared with known HCV gene cloned before the present invention. However, the regions A and B contain less variations, indicating that the polypeptide may be cleaved by signal peptidase at C-terminus of the No. 30 alanine.

The polypeptide is assumed to be envelope protein of HCV or a fragment thereof on the basis of the low homology between the base sequence encoding said polypeptide and a known base sequence which encodes a corresponding region of HCV protein. The polypeptide comprising said 192 amino acid sequence is herein referred to as "M-gp35 protein" or "M-gp35 region".

Polypeptides having 19 amino acid sequence from No. 134 to No.152 of SEQ ID NO 4, 17 amino acid sequence from No. 223 to 239 of SEQ ID NO 4, and 18 amino acid sequence from No. 92 to No.109 of SEQ ID NO 4 are relatively highly hydrophilic and highly homologous to polypeptides having amino acid sequences deduced from known HCV genes and are useful as HCV-associated antigenic peptide in diagnosis and/or for the preparation of vaccine.

Polypeptides having 18 amino acid sequence from No. 40 to No.54 of SEQ ID NO 4, and 12 amino acid sequence from No. 240 to No.251 of SEQ ID NO 4 are relatively highly hydrophilic and extremely low in homology with polypeptides having amino acid sequences deduced from known HCV genes cloned before the present invention and are useful as HCV-associated antigenic peptide for diagnosis.

These polypeptides can be produced by chemical synthesis, as well as by DNA recombinant technique. Furthermore, polypeptides having 76 amino acid sequence from No. 31 to No. 106 of SEQ ID NO 4; and 36 amino acid sequence from No. 134 to No. 169 of SEQ ID NO 4, which correspond to an epitopic region of M-gp35 protein can be produced in larg scale by DNA recombinant technique and are useful as diagnostic reagent and/or vaccine.

5 [2] Gene Encoding NS1(gp70) Region

10

20

50

(1) Preparation of cDNA clone of SEQ ID NO 13 - 27 and sequencing thereof

The cDNA clones of SEQ ID NO 13 - 27 which encode a novel polypeptide of NS1(gp70) region of HCV protein and fragments thereof were cloned from serum from HC patients as follows.

The cloning and sequencing of cDNA encoding HCV polypeptide can be carried out using any of known methods. However, it is hardly accomplished by known "Okayama-Berg" or "Gubler-Hoffman" method because the content of HCV in serum is only a slight amount and HCV gene is liable to variation. The present inventors succeeded in the cloning of gene from a slight amount of serum as will be hereinafter described in Example 1. Briefly, it was conducted by extracting nucleic acids from a serum of a patient suffering from HC. It is preferable to use serum showing OD value of 3.5 on a screening kit of Ortho. Before the extraction, it is desirable to add tRNA or polyribonucleoside to the serum as a carrier for viral RNA. For the purpose of the invention, tRNA is preferable because the degradation of RNA can be easily detected, at least after the addition of tRNA, by monitoring the existence of a sufficient amounts of tRNA having an intact length on electrophoresis.

The resultant RNA is converted into cDNA using transcriptase in the presence of an appropriate oligonucleotide primer. The cDNA is then cloned and amplified by modified PCR (Saiki et al., Nature 324: 126 (1986)) in the presence of a pair of primers. Although commercially available random primers can be used in the above procedures, synthetic primers having the following base sequences are suitable for the present invention.

Synthetic Primers for Cloning of HCV Gene

20

25

30

5'

MS122:GAGGCCGTGAACTGCGATGA(SEQ ID NO:112)

MS148:TTCTCTAAGGTGGCNTCNGCNTG(SEQ ID NO:113)

MS157:CCGGACGCGTTGAANCTNTGNGT(SEQ ID NO:114)

MS123:CATCCAGGTACAACCGAACCA(SEQ ID NO:115)

MS146:AACACACGGCCGCCNCANGGNAA(SEQ ID NO:116)

MS156:CCGGATCCCACAAGCCGTNGTNGA(SEQ ID NO:117)

35

45

In the above sequences, the letter "N" refers to inosine. The above sequences are only illustrative and these base sequences are not critical. They can be modified by replacing nucleotide(s) with other(s), or deleting or inserting nucleotide(s). The replacement may be preferably introduced within 10 bases from 5' terminus involving 1 to several nucleotides, more preferably, within 5 bases involving less than 5 nucleotides. The deletion may occur in the 5' terminal region involving 4 to 5 nucleotides, preferably, within several bases from the 5' terminus involving a few nucleotides. In case of insertion, it may be an addition of 8 to 12, preferably 5 to 6, more preferably, a few nucleotides in 5' terminal region.

PCR can be conducted under appropriate conditions, for example, those described in Example 9 using the first complemental DNA (1st cDNA) as a template. The condition may vary depending on the primers used such as base sequence or combination, length to be amplified, or the like. Examples of pair of primers are: MS122 - MS123; MS157 - MS156; and MS148 - MS146. The resultant cDNA is then inserted into an appropriate site of a cloning vector such as at Smal site of pUC19. A cloning vector harboring the DNA fragment is subjected to the determination of base sequence. Generally, three clones obtained independently are employed and the base sequence of the both strands are determined to obtain an entire base sequence. The sequence is conveniently determined using a fluoresc nce sequencer GENESIS 2000 (DUPONT) according to the protocol attached thereto. Alternatively, a conventional subcloning can be used when the DNA fragment consists more than 180 nucleotides or contains a region which is hardly determined by fluorescence sequencer.

Thus obtained base sequences of DNA fragments are shown in SEQ ID NO 13 to 27.

Clones N19-1, 2, 3, N27-1, 2 and 3 were obtained from serum of a patient N, and clones H19-2, 4, 10, Y19-4, 6 and 7 were obtained independently from patients H, and Y, respectively. Clones MX24-4, 5 and 13

were obtained from a pool comprising sera from multiple patients.

Clones of SEQ ID NO 13 to 15 were obtained using primers MS157 and MS156 represent the sam region of HCV gene designated as N27. Clones of SEQ ID NO 16 to 24 were obtained using prim rs MS122 and MS123 also represent the same region of HCV gen designated as N19, and clones of SEQ ID NO 25 to 27 were obtained using primers MS148 and MS146 also represent the same region of HCV gene designated as MX24. The comparison between base sequence of each clone and that of known HCV gene (Kato et al., Proc. Natl. Acad. Sci. USA, 87: 9524-9528 (1990); and Takamizawa et al., Journal of Virology, 65,3: 1105-1113 (1991)) indicates that clones align on HCV gene in the other of N27, N19 and MX24, from 5' to 3'. As there are overlapping regions between clones, these regions were used to ligate clones each other as will be hereinafter described.

(2) Ligation of Clones of SEQ ID NO 13 to 27

CDNA clones obtained from serum of HC patients shown by sequences of SEQ ID NO 13 to 27 were ligated in the following manners.

1) Ligation of Clones of SEQ ID NO 13 to 15, and Clones of SEQ ID NO 16 to 24

Each clones of SEQ ID NO 13 to 15 was cleaved at Mlul site at Nos. 330 to 335 from 5' terminus of base sequences of SEQ ID NO 13 to 15 and ligated to the Mlul site at No. 71 from 5' terminus of base sequences of SEQ ID NO 16 to 24 by ligation reaction to yield 27 clones having a DNA fragment which comprises, at 5' region, a DNA fragment from clone N27-1, 2 or 3, and at 3' region, a DNA fragment from clone N19-1, 2, 3, clone H19-2, 4, 10, clone Y19-4, 6 or 7. Thus, by the ligation reaction between N27-3 and N19-1, a clone N27N19-1 of SEQ ID NO 28 was obtained.

2) Ligation of Clones of SEQ ID NO 16 to 24, and Clones of SEQ ID NO 25 to 27

Each clones of SEQ ID NO 16 to 24 was ligated to each clones of SEQ ID NO 25 to 27 by PCR. There obtained 54 kinds of clones. Thus, 27 clones have a DNA fragment which encodes either of polypeptides which contain: from N-terminus (amino-terminus) to amino acid No.131, amino acid sequence comprising 131 amino acid residues from N- to C-termini of SEQ ID NO 16 to 24, and from amino acid No. 132 to C-terminus (carboxy-terminus), amino acid sequence from No. 16 to C-terminus of SEQ ID NO 25 to 27. Thus, a clone obtained by the ligation reaction between N19-1 and MX24-4 is the clone N19MX24A-1 of SEQ ID NO 29. The others have a DNA fragment which encodes either of polypeptides which contain: at N-terminal region, amino acid sequence from N-terminus to amino acid No. 116 of SEQ ID NO 16 to 24, and from amino acid No. 117 to C-terminus (carboxy-terminus), amino acid sequence comprising 209 amino acid sequence from N- to C-termini of SEQ ID NO 25 to 27. Thus, a clone obtained by the ligation reaction between N19-1 and MX24-4 is the clone N19MX24B-1 of SEQ ID NO 30.

3) Ligation of Clones of SEQ ID NO 13 to 27

Each clones of SEQ ID NO 13 to 15 was cleaved at Mlul site at base Nos. 330 to 335 from 5' terminus of base sequences of SEQ ID NO 13 to 15 and ligated to the Mlul site at base Nos. 71 to 76 from 5' of the base sequences of the above (2), 2) by ligation reaction to yield clones of N27MX24 series. Thus, a clone obtained by the ligation reaction between N27-3 and N19MX24A-1 is the clone N27MX24A-1 of SEQ ID NO 31 and a clone obtained by the ligation reaction between N27-3 and N19MX24B-1 is the clone N27MX24B-1 of SEQ ID NO 32.

On the basis of the homology between the amino acid sequence of the clone and that reported previously (Kato et al., Proc. Natl. Acad. Sci. USA, 88: 5547-5551 (1991); and Hijikata et al., in: Congress of Association of Japan Molecular Biology, November 29, 1990), the clone N27N19MX24A-1 proved to be the entire region of a gene encoding gp70 polypeptide reported by Kato et al. Thus, polypeptide comprising amino acids from Nos. 46 to 395 of SEQ ID NO 31 and 32 corresponds to th⁻ gp protein presented by Kato et al.

On the other hand, polypeptide comprising amino acids from Nos. 1 to 45 and polypeptide comprising amino acids from No. 46 to the C-terminus of SEQ ID NO 13 to 15 correspond to the C-terminal region of gp 35 polypeptide and N-terminal region of gp70 polypeptide reported by Hijikata et al, respectively. Further, the amino acid sequence from No. 46 to C-terminus of SEQ ID NO 13 to 15 corresponds to a sequence from N-terminus to amino acid No.67 of gp70 reported by Kato et al, and the amino acid

sequence from N- to C-termini of SEQ ID NO 16 to 24 corresponds to a sequence from amino acid Nos. 42 to 172 of gp70 and represents a fragment of gp70 protein presented by Kato et al.

The amino acid No.1 of SEQ ID NO 25 to 27 corresponds to the amino acid No.158 from N-terminus of a sequenc r ported by Hijikata t al., and also th amino acid No. 350 of SEQ ID NO 25 to 27 corresponds to C-terminal amino acids of gp70 reported by Shimotohno et al., and a polypeptide comprising amino acids from Nos. 194 to C-terminus of SEQ ID NO 25 to 27 corresponds to the N-terminal region of non-structural protein of HCV (NS2).

The ligation products prepared in 2) code all or a part of gp70 polypeptide reported by Hijikata et al. For example, the polypeptide from amino acid Nos. 46 to 395 of SEQ ID NO 31 or 32 corresponds to gp70 protein of Hijikata et al. Although a protein expressed from a HCV gene encoding a polypeptide from amino acid Nos. 46 to 395 of SEQ ID NO 31 or 32 is gp70 protein, said expression product is herein referred to as M-gp70, in contrast with gp70 reported by Hijikata et al in: Congress of Association of Japan Molecular Biology, November 29, 1990.

(3) Expression of Polypeptides Encoded by Clones

30

45

DNA fragments obtained in the above 1) and 2) can be used to produce a recombinant HCAg by constructing an expression vector containing DNA encoding a clone, by inserting the DNA into a known expression vector at an appropriate site of the vector, downstream from a promoter, using a well known method per se, and introducing the expression vector harboring the DNA into a host cell such as Escherichia coli cell, yeast cell or the like according to the method known to one of skill, culturing the transformant in a medium under an appropriate condition, and recovering a product from the cultured broth.

The present invention can be accomplished using any expression vectors which have a promoter at an appropriate site to direct the expression of a DNA encoding HCV polypeptide or a fragment thereof. Expression vectors preferably contain promoter, ribosome binding (SD) sequence, gene encoding HCV polypeptide, transcription termination factor, and a regulator gene.

Expression vectors functional in microorganisms such as Escherichia coli, Bacillus subtilis or the like will preferably comprise promoter, ribosome binding (SD) sequence, HCV-associated-protein-encoding gene, transcription termination factor, and a regulator gene.

Examples of promoters include those derived from Escherichia coli or phages such as tryptophane synthetase (trp), lactose operon (lac), λ phage P_L and P_R , T_5 early gene P_{25} , P_{26} promoter and the like. These promoter may have modified or designed sequence for each expression vector such as pac promoter.

Although the SD sequence may be derived from Escherichia coli or phage, a sequence which has been designed to contain a consensus sequence consisting of more than 4 bases, which is complementary to the sequence at the 3' terminal region of 16S ribosome RNA, may also be used.

The transcription termination factor is not essential. However, it is preferable that an expression vector contains a ρ -independent factor such as lipoprotein terminator, trp operon terminator or the like.

Preferably, these sequences required for the expression of a gene encoding HCAg originated from HCV are located, in an appropriate expression plasmid, in the order of promoter, SD sequence, said gene and transcription termination factor from 5' to 3' direction.

Typical example of expression vectors is commercially available pKK233-2 (Pharmacia). However, a series of plasmids pGEX (Pharmacia), which are provided for the expression of fused protein, are also employable for the expression of HCAg-encoding gene of the present invention.

A suitable host cell such as Escherichia coli can be transformed with an expression vector comprising a DNA of the invention by any of known methods such as protocol provided by TOYOBO Japan as described in Example 10.

The cultivation of the transformants can be carried out using any of well known procedures in literatures such as Molecular Cloning, 1982, and the like. The cultivation is preferably conducted at a temperature from about 28°C to 42°C.

Expression vectors used for transforming other host cells, such as those derived from insects, consist of substantially the same elements as those described in the above. However, there are certain preferable factors as follows.

When insect cells are used, a commercially available kit, MAXBAC™ is employed according to the teaching of the supplier (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4). In this case, it is desirable to make a modification to reduce the distance between the promoter of polyhedrin gene and the initiation codon so as to improve the expression of the gene.

A clone of the invention can be inserted into an expression vector for procaryotic cells such as E.coli or

eucaryotic cells such as animal cells after modifying the DNA sequence to bring it in conformity with a fram of initiation codon of said vector. Alternatively, an initiation codon is add d at the 5' terminus of DNA so as to an appropriate translational frame can be produced. The term "a translational frame" of a clone refers to a frame of a bas sequenc in which bases are described as triplets capable of encoding amino acid sequence as illustrated in SEQ ID NO 13 to 32.

The recombinant polypeptide expressed by host cells such as microorganisms including E. coli and insect cells and animal cells can be recovered from the cultured broth by known methods and identified by, for example, immunoreactions between the expressed product and antiserum obtained from HC patients using a conventional method such as Western blot analysis.

E. coli cells transformed with any of clones obtained in the above (1) and (2) express polypeptides encoded thereby as a single polypeptide without cleaving between regions gp35, gp70 and NS2.

When any of clones obtained in the above (1) and (2) is expressed in insect cells, the expressed polypeptide is cleaved between regions gp35, gp70 and NS2. Thus, clone N27MX24A-1 or N27MX24B-1 was transformed into insect or animal cells, polypeptide M-gp70 derived from each clone N27MX24A-1 or N27MX24B-1 was expressed as a glycoprotein after processing.

The following polypeptides comprising amino acid sequence of SEQ ID NO 31 or 32 are relatively highly hydrophilic and homologous to amino acid sequence deduced from known HCV gene cloned before the present invention: a polypeptide consisting of 13 amino acids from amino acid Nos. 143 to 155; a polypeptide consisting of 21 amino acids from amino acid Nos. 171 to 191 subject to that it contains at least amino acids from Nos. 182 to 187; a polypeptide consisting of 14 amino acids from amino acid Nos. 202 to 215 subject to that it contains at least amino acids from Nos. 202 to 209; a polypeptide consisting of 13 amino acids from amino acid Nos. 244 to 256; and a polypeptide consisting of 21 amino acids from amino acid Nos. 299 to 319.

The M-gp70 is a glycoprotein which located adjacent to C-terminus of envelope protein (M-gp35) on HCV gene and contains potential trans-membrane region. These facts lead to an assumption that all or a part of gp70, whose function has not been established yet, may be a part of envelope protein. On the basis of this assumption, the above five kinds of polypeptide fragments, as well as a polypeptide consisting of 106 amino acids from Nos. 109 to 214 and that consisting of 92 amino acids of amino acid sequence from Nos. 233 to 324 of SEQ ID NO 31 or 32, which include said fragments, are useful as vaccine.

Furthermore, the following polypeptides which comprise amino acid sequence of SEQ ID NO 31 or 32 and are expected to be epitopic region of M-gp70 are also useful as vaccine: a polypeptide consisting of 10 amino acids from amino acid Nos. 252 to 261 subject to that it contains at least amino acids from Nos. 252 to 256; a polypeptide consisting of 34 or less than 34 amino acids from amino acid Nos. 250 to 283 subject to that it contains at least amino acids from Nos. 273 to 279; a polypeptide consisting of 20 amino acids from amino acid Nos. 77 to 96; a polypeptide consisting of 18 amino acids from amino acid Nos. 306 to 323; and a polypeptide consisting of 16 amino acids from amino acid Nos. 122 to 137.

The following polypeptides comprising amino acid sequence of SEQ ID NO 31 or 32 are relatively highly hydrophilic and low in homology with amino acid sequences deduced from known HCV genes cloned before the present invention: a polypeptide consisting of 12 amino acids from amino acid Nos. 136 to 147 subject to that it contains at least amino acids from Nos. 136 to 142; a polypeptide consisting of 27 amino acids from amino acid Nos. 45 to 71 subject to that it contains at least amino acids from Nos. 53 to 69; a polypeptide consisting of 9 amino acids from amino acid Nos. 193 to 201.

These polypeptides can be produced by chemical synthesis, as well as by DNA recombinant technique. Furthermore, a polypeptide having 106 amino acid sequence from Nos. 109 to 214 and a polypeptide having 92 amino acid sequence from Nos. 233 to 324 of SEQ ID NO 31 or 32 can be produced in large scale by DNA recombinant technique.

[3] Genes Encoding NS2 - NS4 Regions

(1) Preparation of cDNA clone of SEQ ID NO 33 - 44 and sequencing thereof

The cDNA clones of SEQ ID NO 33 - 44 which encode a novel polypeptide of NS2 - NS4 regions of HCV protein and fragments thereof wer cloned from serum from HC patients as follows.

The cloning and sequencing of cDNA encoding HCV polypeptide can be carried out using any of known methods. However, it is hardly accomplished by known "Okayama-Berg" or "Gubler-Hoffman" method because the content of HCV in serum is only a slight amount and HCV gene is liably to variation. The present inventors succeeded in the cloning of gene from a slight amount of serum as will be hereinafter described in Example 1. Briefly, it was conducted by extracting nucleic acids from a serum of a patient

suffering from HC. It is preferable to use serum showing OD value of 3.5 on a screening kit of Ortho. Before the extraction, it is desirable to add tRNA or polyribonucleoside to the serum as a carrier for viral RNA. For the purpose of the invention, tRNA is pr ferable because the degradation of RNA can be easily detected, at least after the addition of tRNA, by monitoring the existenc of a sufficient amounts of tRNA having an intact length on electrophoresis.

The resultant RNA is converted into cDNA using transcriptase in the presence of an appropriate oligonucleotide primer. The cDNA is then cloned and amplified by modified PCR (Saiki et al., Nature 324: 126 (1986)) in the presence of a pair of primers. Although commercially available random primers can be used in the above PCR, synthetic primers having the following base sequences are suitable for the present invention.

Synthetic Primers for Cloning of HCV Gene

35

40

45

MS49:GACATGCATGTCATGATGTA(SEQ ID NO:118)

MS88:GGCTGCAGCCGGTTCATCCACTGCAC(SEQ ID NO:119)

MS100:GCGGATCCTGCTTCGCCCAGAAGGTC(SEQ ID NO:120)

MS132:GACACATGTGTTGCAGTCGATC(SEQ ID NO:121)

MS152:CGGTCCNAGNAGTATCTCNTTNCC(SEQ ID NO:122)

MS158:ATGGGCCCGGGNGANAGNAGNCTCCCCCTNCTNTC(SEQ ID NO:123)

MS48:GGCTATACCGGCGACTTCGA(SEQ ID NO:124)

MS86:GCGGATCCGGCCTCACCCACATAGATG(SEQ ID NO:125)

MS97:GCGGATCCTCCACCTCCATCGTG(SEQ ID NO:126)

MS135:CTGCTGTCGCCCNGNCCCAT(SEQ ID NO:127)

MS151:ATCACGTGGGGNGCAGANACNGC(SEQ ID NO:128)

MS155:TGTGCCTGNTTNTGGATGATG(SEQ ID NO:129)

In the above sequences, the letter "N" refers to inosine. The above sequences are only illustrative and these base sequences are not critical. They can be modified by replacing nucleotide(s) with other(s), or deleting or inserting nucleotide(s). The replacement may be preferably introduced within 10 bases from 5' terminus involving 1 to several nucleotides, more preferably, within 5 bases involving less than 5 nucleotides. The deletion may occur in the 5' terminal region involving 4 to 5 nucleotides, preferably, within several bases from the 5' terminus involving a few nucleotides. In case of insertion, it may be an addition of 8 to 12, preferably 5 to 6, more preferably, a few nucleotides in 5' terminal region. Primers MS86, MS97, and MS100 contains additional 8 nucleotides encoding a restriction sit at 5' terminus (MS88: 5' GGCTGCAG 3'; MS86, MS97 and MS100: 5' GCGGATCC 3'), however, these are not critical for the isolation of the desired DNA fragments.

PCR can be conducted under appropriate conditions, for example, those described in Example 15 using the first complemental DNA (1st cDNA) as a template. The condition may vary depending on the primers used such as base sequence or combination, length to be amplified, or the like. Examples of pair of primers

are: MS48 - MS49; MS86 - MS100; MS97 - MS88; MS135 - MS132; MS155 - MS152; and MS151 - MS158. The resultant cDNA is then inserted into an appropriate site of a cloning vector such as at Small site of pUC19. A cloning vector harboring the DNA fragment is subjected to the determination of base sequence. Generally, three clon's obtained independently are employed and the base sequence of the both strands are determined to obtain an entire base sequence. The sequence is conveniently determined using a fluorescence sequencer GENESIS 2000 (DUPONT) according to the protocol attached thereto. Alternatively, a conventional subcloning can be used when the DNA fragment consists more than 180 nucleotides or contains a region which is hardly determined by fluorescence sequencer. Thus obtained base sequences of DNA fragments are shown in SEQ ID NO 33 - 39, 44 - 55, and 103 and 104.

Clones N13-1, N15-1, N16 and N23 were obtained from serum of a patient N, clone O26 from patient O, clone U16-4 from patient U, and clone MX25 from a pool comprising sera from multiple patients. Clone of SEQ ID NO 37 obtained from primers MS48 and MS49 and clones of SEQ ID. Nos. 53 to 55 represent the same region on HCV gene (N16 region). The region of clones of SEQ ID NO 44 to 46 obtained using primers MS155 and MS152 on HCV gene was designated as MX25 region. In the same manner, regions of clones of SEQ ID NO 47 to 49, and regions of clones of SEQ ID NO 50 to 52, each obtained by primers MS151 and MS158, or MS135 and MS132, were designated as O26 and N23 regions, respectively.

Clones N13-1, N15-1, O15-1, and O15-2 of SEQ ID NO 38, 39, 103, and 104, which were obtained by primers MS86 and MS100, MS97 and MS88, were designated as regions N13 and N15.

The comparison between base sequences of each clone and known HCV gene (Kato et al., Proc. Natl. Acad. Sci. USA, 87:9524-9528 (1990); and Takamizawa et al., Journal of Virology, 65,3: 1105-1113 (1991)) indicates that clones align in the other of MX25, O26, N23, N16, N13 and N15, from 5' to 3' on the gene.

The clone N16 of SEQ ID NO 36 was obtained by isolating independently three plasmids containing DNA fragment of N16 region, and determining the entire base sequence of DNA fragment originated from HCV.

As there are overlapping region between clones, these regions were used to ligate clones each other.

Clones are highly homologous though, they are distinguishable from each other in terms of nucleotide and amino acid sequences (e.g., clones of SEQ ID NO 33, 34 and 35), which indicates that one patient may carry more than one HCVs at the same time. It is generally accepted that core protein is well conserved even in HCV. When core-protein-encoding gene was cloned in the same manner as that used for the cloning of gene encoding HCV polypeptide, few variations were observed between clones. Among regions on HCV gene, MX25, O26, N23 and N16 regions, especially MX25 region, appear to be highly liable compared with core-protein-encoding region and upperstream region thereof.

(2) Ligation of Clones of SEQ ID NO 33 to 39

10

25

35

40

cDNA clones obtained from serum of HC patients shown by sequences of SEQ ID NO 33 to 37 and 39 were ligated in the following manners.

1) Ligation of Clone N16 of SEQ ID NO 36 and clone N15-1 of SEQ ID NO 39

The ligation was conducted at restriction sites common to both clones. Thus, clone 16 was digested with restriction enzyme to cleave at the BstEll site located at nucleotide Nos. 576-582 of SEQ ID NO 36 and ligated to the BstEll site of clone N15-1 at Nos. 114 to 120 of SEQ ID NO 39 to obtain a DNA fragment consisting of DNA fragments from clones N16 and N15-1 from 5' to 3'. The resultant clones are summarized as clone of SEQ ID NO 41.

2) Ligation of Clones MX25 (SEQ ID NO 33) and O26 (SEQ ID NO 34)

Clones MX25 and O26 were ligated by PCR. By this procedure, multiple DNA fragments encoding different polypeptides were obtained, for example, a DNA fragment encoding a polypeptide which comprises, at the N-terminal region, 284 amino acids of N- to C-termini of SEQ ID NO 33 and, from amino acid No. 285 to the C-terminus, amino acids from No. 32 to the C-terminus of SEQ ID NO 34; a DNA fragment encoding a polypeptide which comprises, at the N-t rminal region, amino acid residues of N-terminus to amino acid No. 252 of SEQ ID NO 33 and, from amino acid No. 253 to the C-terminus, 174 amino acid residues from N- to C- t rmini of SEQ ID NO 34. Thus obtained fused clones were inclusively shown in SEQ ID NO 40.

Clones of SEQ ID NO 36 and 39 or clones of DEQ ID NO 37 and 39 can be ligated by PCR and the resultant clone is shown in SEQ ID NO 41 together with a base sequence obtained in the above 1). Clone

MX25 of SEQ ID NO 33 and clone O26 of SEQ ID NO 34, both of which contain different DNA fragments from those used in the above, were ligated to give multiple DNA fragments having different base sequences. These base sequences are summarized in SEQ ID NO 40.

3) Ligation of Clones of SEQ ID NO 35 and 41

Ligation of clones N23 and N16N15 can be conducted in the same manner as the above 1) to obtain various clones which are designated as N23N15 of SEQ ID NO 42 inclusively. The following illustrative DNA fragments were obtained: a DNA fragment encoding a polypeptide comprising, at the N-terminal region, 307 amino acid residues from N-to C-termini of SEQ ID NO 35 (clone N23) and, from amino acid No. 308 to the C-terminus, amino acids from No.17 to C-terminus of SEQ ID NO 41; a DNA fragment encoding a polypeptide comprising, at the N-terminal region, amino acids from N-terminus to amino acid No. 291 of SEQ ID NO 33 and from amino acid No. 292 to the C-terminus, 477 amino acid residues from N- to C-termini of SEQ ID NO 41.

4) Ligation of Clones of SEQ ID NO 40 and 42

15

20

Ligation of clones MX25O26 and N23N15 can be conducted in the same manner as the above 1) to obtain clones shown in SEQ ID NO 43 inclusively.

The protease activity of viral protein of Flavivirus, a related strain of HCV, exists in the N-terminal domain of non-structural protein of said virus (see, Proc. Natl. Acad. Sci. USA, 87: 8898-8902 (1990)). It is likely that the protease activity of HCV protein also exists in the presumed N-terminal region, NS3. It was confirmed that clone MX25N15 comprises the known entire amino acid sequence encoded by HCV gene (Kato et al., Proc. Natl. Acad. Sci. USA, 87: 9524-9528 (1990)), and a region responsible for the protease activity reported by Hijikata et al (in: Congress of Japan Cancer Association (NIHON Gan-Gakkai (1991)).

Although the both of N- and C-termini of NS3 domain of HCV protein had not been established, it can be presumed to be a region between amino acid Nos. 276 and 884 of SEQ ID NO 43 (clone MX25N15) on the basis of the primary structure of regions to be cleaved by protease and hydrophilic and hydrophobic patterns of Flavivirus protein, referring to a literature (Houghton et al. Hepatology, 14, 2: 381-388 (1991)). The presumed NS3 region of clone MX25N15 is hereinafter referred to as MK/NS3 region.

In the same manner, the NS2 region was presumed to be a polypeptide region between amino acid Nos. 3 and 275 of SEQ ID NO 43 (clone MX25N15) and 40 (clone MX25O26). The presumed NS2 region is hereinafter referred to as MK/NS2 region.

(3) Expression of Polypeptides Encoded by Clones

DNA fragments obtained in the above 1) and 2) can be used to produce a recombinant HCAg by constructing an expression vector containing DNA encoding a clone, by inserting the DNA into a known expression vector at an appropriate site of the vector, downstream from a promoter, using a well known method per se, and introducing the expression vector harboring the DNA into a host cell such as Escherichia coli cell, yeast cell, animal cell or the like according to the method known to one of skill, culturing the transformant in a medium under an appropriate condition, and recovering a product from the cultured broth.

The present invention can be accomplished using any expression vectors which have a promoter at an appropriate site to direct the expression of a DNA encoding HCV polypeptide or a fragment thereof. Expression vectors preferably contain promoter, ribosome binding (SD) sequence, gene encoding HCV polypeptide, transcription termination factor, and a regulator gene.

Expression vectors functional in microorganisms such as Escherichia coli, Bacillus subtilis or the like will preferably comprise promoter, ribosome binding (SD) sequence, HCV-associated-protein-encoding gene, transcription termination factor, and a regulator gene.

Examples of promoters include those derived from Escherichia coli or phages such as tryptophane synthetase (trp), lactose operon (lac), λ phage P_L and P_R , T_5 early gene P_{25} , P_{26} promoter and the like. These promoter may have modified or designed sequenc for each expression vector such as pac promoter.

Although the SD s quence may be derived from Escherichia coli or phage, a s quence which has been designed to contain a consensus sequence consisting of more than 4 bases, which is complementary to the sequence at the 3' terminal region of 16S ribosome RNA, may also be used.

The transcription termination factor is not essential. However, it is preferable that an expression vector

contains a p-independent factor such as lipoprotein terminator, trp operon terminator or the like.

Preferably, these sequences required for the expression of a gene encoding HCAg originated from HCV are located, in an appropriate expression plasmid, in the order of promoter, SD sequence, HCV-associated-protein-encoding gene and transcription t rmination factor from 5' to 3' direction.

Typical example of expression vectors is commercially available pKK233-2 (Pharmacia). However, a series of plasmids pGEX (Pharmacia), which are provided for the expression of fused protein, are also employable for the expression of HCAg-encoding gene of the present invention.

A suitable host cell such as Escherichia coli can be transformed with an expression vector comprising a DNA of the invention by any of known methods such as protocol provided by TOYOBO Japan as described in Example 16.

The cultivation of the transformants can be carried out using any of well known procedures in literatures such as Molecular Cloning, 1982, and the like. The cultivation is preferably conducted at a temperature from about 28°C to 42°C.

Expression vectors used for transforming other host cells, such as those derived from insects or animals including mammals, consist of substantially the same elements as those described in the above. However, there are certain preferable factors as follows.

When insect cells are used, a commercially available kit, MAXBAC™ is employed according to the teaching of the supplier (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4). In this case, it is desirable to make a modification to reduce the distance between the promoter of polyhedrin gene and the initiation codon so as to improve the expression of the gene.

When animal cells are used as hosts, expression vectors preferably contain active-type promotor from adenovirus EIA gene (ZOKUSEIKAGAKU JIKKEN KOZA I, IDENSHI KENKYU-HO II, 189-190, 1986), SV40 early promoter, SV40 late promoter, apolipoprotein E gene promoter, SRα promotor (Molecular and Cellular Biology, 8, 1, 466-472, 1988) or the like. Specifically, known expression vectors such as pKCR (Proc. Natl. Acad. Sci. USA, 78: 1528 (1981)) or a derivative thereof, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), which prepared by modifying pKCR maintaining its essential functions, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), or the like may be employed.

Animal cells usable in the present invention are CHO cell, COS cell, mouse L cell, mouse C127 cell, mouse FM3A cell and the like.

A clone of the invention can be inserted into an expression vector for procaryotic cells such as <u>E.coli</u> or eucaryotic cells such as animal cells after modifying the DNA sequence to bring it in conformity with a frame of initiation codon of said vector. Alternatively, an initiation codon is added at the 5' terminus of DNA so as to an appropriate translational frame can be produced. The term "a translational frame" of a clone refers to a frame of a base sequence in which bases are described as triplets capable of encoding amino acid sequence as illustrated in SEQ ID NO 33 to 43.

The recombinant polypeptide expressed by the host cells such as microorganisms including E. coli, insect cells and animal cells can be recovered from the cultured broth by known methods and identified by, for example, immunoreactions between the expressed product and antiserum obtained from HC patients using a conventional method such as Western blot analysis.

Hydrophilic study and prediction of higher-order structure of protein, the following peptide fragments contained in a polypeptide having amino acid sequence of SEQ ID NO 43 appeared to be highly hydrophilic and can take so-called "turn structure" not α -helix or β -sheet structure in high probability. Therefore, these fragments possibly represent antigen determinants, or can contain at least one antigen determinant of HCAg. Although the higher-order structure in serum and the specific reactivity of each fragment are not established, it can be concluded that the following peptide fragments are highly reactive with antiserum raised against HCV-associated antigens. A polypeptide consisting of 19 amino acids from amino acid Nos. 247 to 265 of SEQ ID NO 43; a polypeptide consisting of 8 to 25 amino acids subject to that it contains at least 8 amino acids from Nos. 300 to 307; a polypeptide consisting of 13 to 25 amino acids subject to that it contains at least 13 amino acids from Nos. 410 to 428; a polypeptide consisting of 10 amino acids from Nos. 283 to 292; a polypeptide consisting of 14 amino acids from Nos. 477 to 490; a polypeptide consisting of 14 amino acids from Nos. 498 to 512; a polypeptide consisting of 12 amino acids from Nos. 538 to 550; a polypeptide consisting of at least 21 amino acids from Nos. 747 to 767; a polypeptide consisting of at least 12 amino acids from Nos. 841 to 852; a polypeptide consisting of at least 12 amino acids from Nos. 867 to 878; a polypeptide consisting of 8 to 25 amino acids subject to that it contains at least 8 amino acids from Nos. 665 to 672; and a polypeptide consisting of 15 amino acids from Nos. 315 to 327.

The above polypeptide fragments can be obtained by means of chemical synthesis, as well as DNA recombinant technique.

Other polypeptide fragments of clone of SEQ ID NO 43, that is, a polypetides containing the entire or a

part of a polypeptide consisting of 266 amino acids from Nos. 461 to 726; a polypeptide consisting of 74 amino acids from Nos. 477 to 550; a polypeptide consisting of 42 amino acids from Nos. 963 to 1004; and a polypeptide consisting of 45 amino acids from Nos. 283 to 327, can be prepared in a large scale by recombinant DNA technique.

[4] Gene Encoding NS4 to NS5 Regions

(1) Preparation of cDNA clone of SEQ ID NO 64 - 75 and sequencing thereof

The cDNA clones of SEQ ID NO 64 - 75 which encode a novel polypeptide of NS4 to NS5 regions of HCV protein and fragments thereof were cloned from serum from HC patients as follows.

The cloning and sequencing of cDNA encoding HCV polypeptide can be carried out using any of known methods. However, it is hardly accomplished by known "Okayama-Berg" or "Gubler-Hoffman" method because the content of HCV in serum is only a slight amount and HCV gene is liable to variation. The present inventors succeeded in the cloning of gene from a slight amount of serum as will be hereinafter described in Example 1. Briefly, it was conducted by extracting nucleic acids from a serum of a patient suffering from HC. It is preferable to use serum showing OD value of 3.5 on a screening kit of Ortho. Before the extraction, it is desirable to add tRNA or polyribonucleoside to the serum as a carrier for viral RNA. For the purpose of the invention, tRNA is preferable because the degradation of RNA can be easily detected, at least after the addition of tRNA, by monitoring the existence of a sufficient amounts of tRNA having an intact length on electrophoresis.

The resultant RNA is converted into cDNA using transcriptase in the presence of an appropriate oligonucleotide primer. The cDNA is then cloned and amplified by means of polymerase chain reaction (PCR) (Saiki et al., Nature 324: 126 (1986)) in the presence of a pair of primers. Although commercially available random primers can be used in the PCR, synthetic primers having the following base sequences are suitable for the present invention.

Synthetic Primers for Cloning of HCV Gene

30

35

40

45

50

55

5'

MS126:GGTGAGCATGGAGGTGACCAC(SEQ ID NO:130)

MS119:TCATCCTCCTCCGCTCGAAGC(SEQ ID NO:131)

MS161:GTGGACGCCTTNGCCTTCATNTC(SEQ ID NO:132)

MS162:ACGGATGTCNTTCTCNGTNAC(SEQ ID NO:133)

MS121:GGCGGAATTCCTGGTCATAGCCTCCGTGAA(SEQ ID NO:134)

MS163:GGGGNATGGCCTATTGGCCTG(SEQ ID NO:135)

MS127:GGCATGTGGGCCCAGGGGAGG(SEQ ID NO:136)

MS118:TGTGAGCCCGAACCGGATGT(SEQ ID NO:137)

MS159:GTGGTANTCCTGGACTCNTTNGA(SEQ ID NO:138)

MS160:ACTACCGNGACGTGCTNAANGA(SEQ ID NO:139)

MS120:TGGGGATCCCGTATGATACCCGCTGCTTTG(SEQ ID NO:140)

MS174:ATTGTCAGATCTACGGGGCCACTT(SEQ ID NO:141)

MS175:GCAAGCTTAAAAAAAAAAAAGGGGGGATGGCCTATTGGCCTGGA(SEQ ID

30 NO:142)

5

15

25

In the above sequences, the letter "N" refers to inosine. The above sequences are only illustrative and these base sequences are not critical. They can be modified by replacing nucleotide(s) with other(s), or deleting or inserting nucleotide(s). The replacement may be preferably introduced within 10 bases from 5' terminus involving 1 to several nucleotides, more preferably, within 5 bases involving less than 5 nucleotides. The deletion may occur in the 5' terminal region involving 4 to 5 nucleotides, preferably, within several bases from the 5' terminus involving a few nucleotides. In case of insertion, it may be an addition of 8 to 12, preferably 5 to 6, more preferably, a few nucleotides in 5' terminal region.

PCR can be conducted under appropriate conditions, for example, those described in Example 21 using the first complemental DNA (1st cDNA) as a template. The condition may vary depending on the primers used such as base sequence or combination, length to be amplified, or the like. Examples of pair of primers are: MS127 - MS126; MS118 - MS119; MS159 - MS161; MS160 - MS162; MS120 - MS163; and MS120 - MS121. The resultant cDNA is then inserted into an appropriate site of a cloning vector such as at Smal site of pUC19. A cloning vector harboring the DNA fragment is subjected to the determination of base sequence. Generally, three clones obtained independently are employed and the base sequence of the both strands are determined to obtain an entire base sequence. The sequence is conveniently determined using a fluorescence sequencer GENESIS 2000 (DUPONT) according to the protocol attached thereto. Alternatively, a conventional subcloning can be used when the DNA fragment consists more than 180 nucleotides or contains a region which is hardly determined by fluorescence sequencer. Thus obtained base sequences of DNA fragments are shown in SEQ ID NO 64 - 69, and 76 - 100.

Clones N22-1, 3, N17-1, 2, 3, N29-1, 2, 3, N18-2, 3 and 4 wer obtained from serum of a patient N, clone H22-3, 8, 9, H17-1, 3, H18-1, 2 and 3 from patient H, clone O28-1, 2, 4, O30-2, 3 and 4 from patient O. It is generally accepted that region encoding core protein or its 5' region generally contain few variations and are well conserved even in HCV. When regions encoding core protein and/or a upstream region thereof were cloned in the same manner as the above, variations were hardly observed between clones. In the present invention, clones obtained from a same region on HCV gene were highly homologous though, they

proved to be DNA fragments distinguishable from each other in terms of nucleotide and amino acid sequences. This indicates that one patient may carry more than one HCVs at the same time.

From the above fact, N22, N17, O28, N18, N29, and O30 regions assumed to be highly liable compared with core-protein-encoding region and upstream region thereof.

Region on HCV gene which corresponds to each clone was designated as follows. The region of clones N22-1, 3, H-22, 3, 8 and 9 obtained using primers MS127 and MS126 was designated as N22. In the same manner, the regions on HCV gene corresponding to clones N17-1, 2, 3, H17-1 and 3 obtained using primers MS118 and MS119, clones O28-1, 2 and 4 obtained using primers MS159 and MS161, clones N29-1, 2 and 3 obtained using primers MS160 and MS162, clones N18-2, 3, 4, H18-1, 2 and 3 obtained using primers MS120 and MS121, clones O30-2, 3 and 4 obtained using primers MS120 and MS163 were designated as N17, O28, N29, N18 and O30, respectively.

The comparison between base sequences of each clone and known HCV gene (Kato et al., Proc. Natl. Acad. Sci. USA, <u>87</u>:9524-9528 (1990); and Takamizawa et al., Journal of Virology, 65,3: 1105-1113 (1991)) indicates that clones align in the other of N22, N17, O28, N29, and O30, from 5' to 3' on the gene (N18 is included in O30 region).

There are overlapping region between clones, which were used to ligate clones each other.

(2) Ligation of Clones of SEQ ID NO 64 to 69

Regions N22, N17, O28, N29, and O30 of clone N15 (see, the above [3]), a cDNA clone obtained from serum of HC patients, were ligated in the following manners.

1) Ligation of N17 and O28 Regions

The ligation of N17 and O28 regions can be conducted using, for instance, clones N17-3 (SEQ ID NO 81) and O28-1 (SEQ ID NO 86). The ligation was carried out by PCR. Thus, about equimolar of DNA fragments (as template) of clones N17-3 and O28-1 in a solution were subjected to PCR in the presence of primers MS118 and MS161 to yield clone 1728.

30 2) Ligation of N29 and N18 Regions

25

50

55

In the same manner as the above 1), N29 and N18 regions were ligated using clones N29-1 (SEQ ID NO 89) and N18-4 (SEQ ID NO 92), and primers MS160 and MS121 to yield clone 2918.

35 3) Ligation of Regions N 17 to N18

PCR was carried out using DNA fragments of clones 1728 and 2918, primers MS118 and MS121 to yield clone 1718 which contains clones N17, O28, N29, N18 from 5' to 3'. The clone 1718 was cloned into Smal site of PUC19 to give plasmid 1718 in which EcoRI site from pUC19, clone N17-3 and N18 regions on HCV gene are aligned in this order from 5' to 3'.

4) Ligation of Regions N 22 to N17

In the same manner as the above 1), DNA fragments of clones N22-1 (SEQ ID NO 76) and N17-3 (SEQ ID NO 81) were ligated by PCR using primers MS127 and MS119 to yield a DNA fragment designated as clone 2217 which contains N22 and N17 from 5' to 3'. The clone 2217 was cloned into Smal site of pUC19 in the same manner as the above 3) to give plasmid 2217 in which EcoRI site located at 5' terminus.

5) Ligation of Clones 2217 and 1718

Upon digestion with restriction

Upon digestion with restriction enzyme Xbal, clone 1718 is cleaved at one site. Plasmid pUC1718 was digested with Xbal and a DNA fragment comprising DNA fragment of clone 1718 and Xbal sit- of pUC19 was isolated. The DNA fragm int derivid from clone 2217 was inserted into Xbal site of pUC2217 such that the Xbal site in N17 region of pUC2217 and Xbal site from pUC19 are ligated to obtain plasmid pUC2218.

6) Ligation of N15 Region and O30 Region Corresponding to 3' Terminal Region of HCV Gene

An example of DNA fragment of O30 region is clone O30-3 of SEQ ID NO 98. Plasmid pUCO30

contains the DNA fragment of O30-3 at Smal site of pUC19 in the order of, from 5' to 3', EcoRI site and clone O30-3. Plasmid pUCN15 contains a DNA fragment of HCV gene, clone N15 (see, [3]), forwardly at Smal site of pUC19 in the order of, from 5' to 3', EcoRI site and clone N15.

Plasmid pUCO30 was cleaved by SacI and blunt ended, which was followed by the cleavage at another cloning site, HindIII, to isolate a DNA fragment from plasmid pUCN15 which was digested with XbaI, blunt ended, digested with HindIII to yield plasmid pUC15-30. Taking advantage of the fact that said plasmid pUC15-30 has only one site which can be cleaved by restriction enzymes BgIII and HindIII, it was subjected to PCR using a primer MS174 having a BgIII site in sequence derived from clone O30-3 in order to add poly U at 3' terminus of clone O30-3.

PCR was conducted using, as a template, pUC15-30 and primers MS174 and MS175. PCR fragment was then digested with Bglll and Hindlll and the resultant fragment ligated to a Bglll-Hindlll fragment of pUCO30 containing the vector fragment of pUCO30 to obtain plasmid pUC15-30U having polyU attached to the 3' terminus of clone O30-3.

7) Ligation of N15 to O30 Regions

20

30

There is an Apal site within a region common to N15 and N22 regions. There is an Apal site within a region common to N18 and O30. A DNA fragment isolated from pUC2218 with Apal was inserted into Apal site of pUC15-30U appropriately to obtain plasmid pUC1530U.

The ligated N15 to O30 regions encodes amino acid sequence which is highly homologous to amino acid sequence of NS5, a part of non-structural protein NS4 of Flavivirus, a related strain of HCV. It was also confirmed that said region is homologous to a sequence encoding a part of NS4 region AND NS5 region by comparison with a known sequence of HCV gene disclosed by aforementioned Chiron, Shimotohno, or Takamizawa. As a conclusion, clone disclosed in Seq. Lis. represents DNA sequence assumed to be NS4 and NS5 regions of HCV gene. The clone was then inserted into an expression plasmid to produce polypeptide encoded by said clone. The polypeptide was then evaluated as to the ability to react immunologically with antiserum of HC patients.

(3) Expression of Polypeptides Encoded by Clones

DNA fragments obtained in the above (2) can be used to produce a recombinant HCAg by constructing an expression vector containing DNA encoding a clone, by inserting the DNA into a known expression vector at an appropriate site of the vector, downstream from a promoter, using a well known method per se, and introducing the expression vector harboring the DNA into a host cell such as Escherichia coli cell, yeast cell, animal cell or the like according to the method known to one of skill, culturing the transformant in a medium under an appropriate condition, and recovering a product from the cultured broth.

The present invention can be accomplished using any expression vectors which have a promoter at an appropriate site to direct the expression of a DNA encoding HCV polypeptide or a fragment thereof. Expression vectors preferably contain promoter, ribosome binding (SD) sequence, gene encoding HCV polypeptide, transcription termination factor, and a regulator gene.

Expression vectors functional in microorganisms such as Escherichia coli, Bacillus subtilis or the like will preferably comprise promoter, ribosome binding (SD) sequence, HCV-associated-protein-encoding gene, transcription termination factor, and a regulator gene.

Examples of promoters include those derived from Escherichia coli or phages such as tryptophane synthetase (trp), lactose operon (lac), λ phage P_L and P_R , T_5 early gene P_{25} , P_{26} promoter and the like. These promoter may have modified or designed sequence for each expression vector such as pac promoter.

Although the SD sequence may be derived from Escherichia coli or phage, a sequence which has been designed to contain a consensus sequence consisting of more than 4 bases, which is complementary to the sequence at the 3' terminal region of 16S ribosome RNA, may also be used.

The transcription termination factor is not essential. However, it is preferable that an expression vector contains a ρ -independent factor such as lipoprotein terminator, trp operon terminator or the like.

Preferably, these sequences required for the xpression of a gene encoding HCAg originated from HCV are located, in an appropriate expression plasmid, in the order of promoter, SD sequence, said gene and transcription termination factor from 5' to 3' direction.

Typical example of expression vectors is commercially availabl pKK233-2 (Pharmacia). However, a series of plasmids pGEX (Pharmacia), which are provided for the expression of fused protein, are also employable for the expression of HCAg-encoding gene of the present invention.

A suitable host cell such as Esch richia coli can b transformed with an expression vector comprising a DNA of the invention by any of known methods such as protocol provided by TOYOBO Japan as described in Example 22.

The cultivation of the transformants can b carried out using any of well known proc dur s in literatur s such as Molecular Cloning, 1982, and the like. The cultivation is preferably conducted at a temperature from about 28°C to 42°C.

Expression vectors used for transforming other host cells, such as those derived from insects or animals including mammals, consist of substantially the same elements as those described in the above. However, there are certain preferable factors as follows.

When insect cells are used, a commercially available kit, MAXBAC™ is employed according to the teaching of the supplier (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4). In this case, it is desirable to make a modification to reduce the distance between the promoter of polyhedrin gene and the initiation codon so as to improve the expression of the gene.

When animal cells are used as hosts, expression vectors preferably contain active-type promotor from adenovirus EIA gene (ZOKUSEIKAGAKU JIKKEN KOZA I, IDENSHI KENKYU-HO II, 189-190, 1986), SV40 early promoter, SV40 late promoter, apolipoprotein E gene promoter, SRα promotor (Molecular and Cellular Biology, 8, 1, 466-472, 1988) or the like. Specifically, known expression vectors such as pKCR (Proc. Natl. Acad. Sci. USA, 78: 1528 (1981)) or a derivative thereof, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), which prepared by modifying pKCR maintaining its essential functions, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), or the like may be employed.

Animal cells usable in the present invention are CHO cell, COS cell, mouse L cell, mouse C127 cell, mouse FM3A cell and the like.

A clone of the invention can be inserted into an expression vector for procaryotic cells such as E.coli or eucaryotic cells such as animal cells after modifying the DNA sequence to bring it in conformity with a frame of initiation codon of said vector. Alternatively, an initiation codon is added at the 5' terminus of DNA so as to an appropriate translational frame can be produced. The term "a translational frame" of a clone refers to a frame of a base sequence in which bases are described as triplets capable of encoding amino acid sequence as illustrated in SEQ ID NO 64 to 75.

The recombinant polypeptide expressed by host cells such as microorganisms including E. coli, insect cells and animal cells can be recovered from the cultured broth by known methods and identified by, for example, immunoreactions between the expressed product and antiserum obtained from HC patients using a conventional method such as Western blot analysis.

Hydrophilic study and prediction of higher-order structure of protein, the following peptide fragments contained in a polypeptide having amino acid sequence of SEQ ID NO 75 appeared to be highly hydrophilic and can take so-called "turn structure" not α -helix or β -sheet structure in high probability. Therefore, these fragments possibly represent antigen determinants, or can contain at least one antigen determinant of HCAg. Although the higher-order structure in serum and the specific reactivity of each fragment are not established, it can be concluded that the following peptide fragments are highly reactive with antiserum raised against HCV-associated antigens. A polypeptide comprising at least 20 amino acids from amino acid Nos. 324 to 343; a polypeptide comprising at least 14 amino acids from Nos. 356 to 369; a polypeptide cornprising at least 18 amino acids from Nos. 584 to 601; a polypeptide comprising 10 amino acids from Nos. 588 to 597; a polypeptide consisting of 10 amino acids from Nos. 620 to 629; a polypeptide consisting of 18 amino acids from Nos. 901 to 918; and a polypeptide which contains at least any of those described in the above and comprises 25 or less amino acids of SEQ ID NO 75.

35

45

The above polypeptide fragments can be obtained by means of chemical synthesis, as well as DNA recombinant technique.

Other polypeptide fragments of SEQ ID NO 75, that is, a polypeptides containing the entire or a part of a polypeptide consisting of 74 amino acids from Nos. 413 to 486; a polypeptide consisting of 997 amino acids from Nos. 415 to 1411; a polypeptide consisting of 74 amino acids from Nos. 655 to 728; a polypeptide consisting of 98 amino acids from Nos. 858 to 955; a polypeptide consisting of 92 amino acids from Nos. 1009 to 1100; a polypeptide consisting of 66 amino acids from Nos. 1160 to 1225; and a polypeptide consisting of 54 amino acids from Nos. 763 to 816 can be prepared in a large scale by recombinant DNA technique.

[5] Pr paration of a cDNA Clone T7N1-30U Originated from Serum of HC Patient (SEQ ID NO 101)

The gene or a DNA fragment encoding a novel polypeptide of SEQ ID NO 101 can be obtained by following procedures.

The ligation of clones N19MX24A-1 and MX25-1 by PCR gives a DNA fragment in which either of the 3' sequence of MX24 region and 5' sequence of MX25 region, which are overlapping each other, is preferentially used (Clone 1925). A synthetic DNA was synthesized in order to introduce into clone N1-1, from 5' to 3', restriction sites HindIII and Spel and T7 promoter and clone T7N1-1 was obtained by cassette ligation. Clone T7N1N3N10 was obtain d in the same manner as that used for the preparation of clone N1N3N10 except that clone T7N1-1 was used instead of clone N1-1. This clone was ligated to clone N27N19-1 by restriction enzyme BamHI to obtain clone T7N119. The clones T7N119 and 1925 have N19 regions and the both clones were ligated using Pvul restriction site in the N19 region to yield clone T7N1-25.

A EcoRI-NotI-BamHI adapter (Toyobo) was ligated to plasmid pUC1530U at the HindIII site in its 3' terminal region to obtain Clone 1530UNot which contains NotI site at 3' terminus of clone 1530U.

10

For the ligation of clones T7N1-25, 1530UNot, and MX25N15-1, prepared in [3], the three clones were ligated at Pstl site in MX25 region common to clones T7N1-25 and MX25N15-1 and EcoT22l site in N15 region common to clones 1530UNot and MX25N15-1. Clone T7N1-25 has Spel site at 5' terminus and clone 1530UNot has Notl site at 3' terminus.

HCV gene can be prepared by ligating clones T7N1-25, MX25N15-1 and 1530UNot in this order without overlapping. Thus, clone T7N1-25 is digested with SpecI and PstI, clone MX25N15-1 with PstI and EcoT221, clone 1530UNot with EcoT22I and NotI, λZapII (Strategene) with SpecI and NotI, respectively, and the resultant fragments were ligated to yield a phage in which a single DNA fragment having a sequence of HCV gene between SpeI and NotI sites of λZapII (from 5' to 3': clone T7N1-25, MX25N15-1 and 1530UNot). The resultant HCV derived clone was designated as T7NI-30U. Ligation to λZapII (Strategene), isolation of phage DNA, subcloning into pBluescriptII can be conducted according to the protocol attached to the kit. The packaging for the preparation of phage particles were carried out using Gigapack II Packaging Extracts (Strategene) according to the protocol attached thereto. The clone T7N1-30U is a DNA fragment which comprises a cDNA originated from HCV having an inserted T7 phage promoter at 5' terminus, and poly T at 3' terminus.

[6] Expression of Fused Polypeptides Encoded by cDNA Originated from Serum of HC Patients

Recombinant HCV-associated antigen can be obtained by expressing all or a part of clones prepared in [1], [2], [3] or [4], or DNA sequence encoding all the protein of HCV prepared in [5].

The present invention can be accomplished using any expression vectors which have a promoter at an appropriate site to direct the expression of a DNA encoding HCV polypeptide or a fragment thereof. Expression vectors preferably contain promoter, ribosome binding (SD) sequence, gene encoding HCV polypeptide, transcription termination factor, and a regulator gene.

Expression vectors functional in microorganisms such as Escherichia coli, Bacillus subtilis or the like will preferably comprise promoter, ribosome binding (SD) sequence, HCV-associated-protein-encoding gene, transcription termination factor, and a regulator gene.

Examples of promoters include those derived from Escherichia coli or phages such as tryptophane synthetase (trp), lactose operon (lac), λ phage P_L and P_R , T_5 early gene P_{25} , P_{26} promoter and the like. These promoter may have modified or designed sequence for each expression vector such as pac promoter.

Although the SD sequence may be derived from Escherichia coli or phage, a sequence which has been designed to contain a consensus sequence consisting of more than 4 bases, which is complementary to the sequence at the 3' terminal region of 16S ribosome RNA, may also be used.

The transcription termination factor is not essential. However, it is preferable that an expression vector contains a ρ -independent factor such as lipoprotein terminator, trp operon terminator or the like.

Preferably, these sequences required for the expression of a gene encoding HCAg originated from HCV are located, in an appropriate expression plasmid, in the order of promoter, SD sequence, said gene and transcription termination factor from 5' to 3' direction.

Typical example of expression vectors is commercially available pKK233-2 (Pharmacia). However, a series of plasmids pGEX (Pharmacia), which are provided for the expression of fused protein, are also employable for the expression of HCAg-encoding gen of the present invention.

A suitable host cell such as Escherichia coli can be transformed with an expression vector comprising a DNA of the invention by any of known methods such as protocol provided by TOYOBO Japan as described in Example 30.

The cultivation of the transformants can be carried out using any of well known procedures in literatures such as Molecular Cloning, 1982, and the like. The cultivation is preferably conducted at a temperature from

about 28°C to 42°C.

50

Expression vectors used for transforming oth r host cells, such as those derived from insects or animals including mammals, consist of substantially the same elements as those described in the above. However, there are certain pr ferable factors as follows.

When insect cells are used, a commercially available kit, MAXBAC™ is employed according to the teaching of the supplier (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4). In this case, it is desirable to make a modification to reduce the distance between the promoter of polyhedrin gene and the initiation codon so as to improve the expression of the gene.

When animal cells are used as hosts, expression vectors preferably contain active-type promotor from adenovirus EIA gene (ZOKUSEIKAGAKU JIKKEN KOZA I, IDENSHI KENKYU-HO II, 189-190, 1986), SV40 early promoter, SV40 late promoter, apolipoprotein E gene promoter, SRα promotor (Molecular and Cellular Biology, 8, 1, 466-472, 1988) or the like. Specifically, known expression vectors such as pKCR (Proc. Natl. Acad. Sci. USA, 78: 1528 (1981)) or a derivative thereof, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), which prepared by modifying pKCR maintaining its essential functions, pBPV MT1 (Proc. Natl. Acad. Sci. USA, 80: 398 (1983)), or the like may be employed.

Animal cells usable in the present invention are CHO cell, COS cell, mouse L cell, mouse C127 cell, mouse FM3A cell and the like.

A clone of the invention can be inserted into an expression vector for procaryotic cells such as E.coli or eucaryotic cells such as animal cells after modifying the DNA sequence to bring it in conformity with a frame of initiation codon of said vector. Alternatively, an initiation codon is added at the 5' terminus of DNA so as to an appropriate translational frame can be produced. The term "a translational frame" of a clone refers to a frame of a base sequence in which bases are described as triplets capable of encoding amino acid sequence as illustrated in SEQ ID NO 1 to 104.

The recombinant polypeptide expressed by host cells such as microorganisms including E. coli, insect cells and animal cells can be recovered from the cultured broth by known methods and identified by, for example, immunoreactions between the expressed product and antiserum obtained from HC patients using a conventional method such as Western blot analysis.

Polypeptide encoded by gene of the invention contains region(s) which seem to be immunologically highly reactive with antiserum of HC. These regions were ligated and expressed in various cells as fused protein. For example, polypeptide having amino acids from Nos. 1 to 115 of SEQ ID NO 3 was expressed using expression vector pCZCORE. The expression vector was modified to replace the 3' region from the epitopic region of said polypeptide with clone N23 which encodes a desired polypeptide to express a fused protein. It was followed by the ligation of a polypeptide having amino acids from Nos. 963 to 1005 of SEQ ID NO 43 to the C-terminus of polypeptide encoded by N23 region. Thus, regions encoding polypeptides which seem to be immunologically highly reactive with antiserum of HC patients were ligated to cDNA and inserted into an expression vector to express said polypeptides.

Specifically, as shown in Example 30, a polypeptide CN23 which contains an epitopic region of core protein of HCV and a region comprising an epitope which is encoded by clone N23, a part of non-structural protein region NS3 and is seem to be immunologically highly reactive with antiserum of HC patients, was expressed directly in E. coli

Thus, clone N23, from No. 107 (G), was inserted in frame into pCZCORE at the SacII site within core gene. Expression vector pCZCN23 capable of expressing epitopic regions of core protein and a polypeptide encoded by N23 as a fused protein was constructed by ligating a part of N23 to the 3' terminus of the N-terminal gene of core protein. A DNA fragment which encodes HCV protein and has SD sequence at 5' terminus was ligated in tandem to the vector, resulting in the expression of desired polypeptide in large scale.

The resultant fused protein comprising epitopic regions of core protein and N23 region reacted with antiserum of HC patient in high probability.

Thus, the present invention provides a novel gene of HCV or a fragment thereof and polypeptide encoded by the same. The recombinant polypeptide is highly reactive with HCAb and can be used for the development of a method for detecting HCAb efficiently, and for the preparation of vaccine. DNA and polypeptides of the invention are also useful for the development of in vivo or in vitro system for the estimation of protease activity of HCV.

The following Examples further illustrate and detail the invention disclosed, but should not be construed to limit the invention. Throughout the Examples concerning the isolation of RNA and cloning of cDNA, tip or pipet used for the preparation of samples and/or reagents employed for reaction was changed to cleaned and/or sterilized one every time for preventing the sample from contamination. The procedures which are not specifically described were conducted substantially in accordance with the teachings of literatures given

in parentheses.

Electrophoresis of nucleic acids (Molecular Cloning (1982), Cold Spring Harbor): cleavage of DNA fragment with restriction enzymes (Molecular Cloning (1982), Cold Spring Harbor); or a catalogue "IDENSHIKOGAKU KENKYU-YO SIYAKU SOGO KATALIOGU", Toyobo): ligation reaction of DNA fragments (TAKARA Biotechnology Catolog, 1991, vol. 1, Takara Shuzo): extraction of DNA from acrylamide gel or agarose gel (Molecular Cloning (1982), Cold Spring Harbor): cultivation of E. coli transformants transformed with a plasmid on agarose plate and isolation of colony therefrom (Molecular Cloning (1982), Cold Spring Harbor).

6 Example 1

Extraction of Nucleic Acids from Serum of a Patient Suffering from Hepatitis C

To 10 ml of a serum from a patient of HC (OD = 3.5 or more on HCV EIA kit of Ortho & Co.) was added 25 ml of Tris buffer (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 100 mM NaCl), mixed and centrifuged (20,000 x g, 20 min) at 20 °C. The supernatant was centrifuged (100,000 x g, 5 hr) at 20 °C. The pellet was dissolved into 1.5 ml of Protenase K solution (1% sodium dodecyl sulfate, 10 mM EDTA, 10 mM Tris-HCl (pH 7.5), 2 mg/ml Protenase K (Pharmacia), and 6.6 µg yeast tRNA mixture) and the solution incubated at 45 °Cfor 90 min. The solution was then subjected to the phenol/chloroform extraction (more than 4 times) which was carried out by adding an equal volume of phenol/chloroform to the solution, vigorously mixing, and centrifuging to recover the aqueous layer containing nucleic acids. It was followed by chloroform treatments (more than two times) and ethanol precipitation. The ethanol precipitation was carried out by mixing the aqueous solution with 2.5 volumes of ethanol containing either of 1/10 volume of 3M sodium acetate or an equal volume of 4 M arnmonium acetate, allowing to stand for overnight at -20 °C, or more than 15 min at -80 °C, centrifuging (35,000 rpm, 4 hr) by SW41 Ti Rotor (Beckman) to pellet nucleic acids, and recovering the pellet. The pellet of nucleic acid was then dried for the subsequent use.

Example 2

Synthesis of cDNA

[1] Preparation of RNA Sample Solution

RNA sample solution was prepared by resolving the dried nucleic acid obtained in Example 1 in 30 µl of water containing 10 µl of ribonuclease inhibitor (100 U/µl, Takara Shuzo, Japan).

[2] Synthesis of cDNA Using Random Primer

To 2 μ I of RNA sample solution was added 2.7 μ I of random primer (0.170D, Amersham), 2 μ I of 10 x PCR (Mg) buffer (100 mM Tris-HCI (pH 8.3), 500 mM KCI, 60 mM MgCl₂), 8 μ I of 1.25 mM 4dNTPs, 2 μ I of water and the mixture incubated at 65 °C for 5 min then at 25 °C for 5 min. To the mixture was added I μ I of reverse transcriptase (25 U, Life Science), 1 μ I of ribonuclease inhibitor (100 U/ μ I, Takara Shuzo) and the mixture incubated at 37 °C for 20 min, at 42 °C for 30 min, and finally at 95 °C for 5 min, which was followed by prompt cooling to 0 °C (synthesis of cDNA).

Amplification of DNA having specific sequences was conducted substantial in accordance with the polymerase chain reaction (PCR) of Saiki et al. (Nature 324: 126 (1986)). Throughout the specification, the expression that PCR was carried out according to Saiki's method means that the PCR was conducted substantial in accordance with the polymerase chain reaction (PCR) of Saiki et al. For the PCR, a 100 μl of a mixture containing 2 μl of cDNA solution, 10 μl of 10 x PCR buffer (100 mM Tris-HCl (pH 8.3), 500 mM KCl, 150 mM MgCl₂, 1% gelatin), 8 μl of 2.5 mM 4 dNTPs, 50 pmol each of two synthetic primers (the pair of primers consists of S1 - AS1, S2 - AS1, S2 - AS2, or S4 - AS3) and water was incubated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, it was mixed with 0.5 μl of Taq DNA polymerase (7 U/μl, AmpliTaqTM, Takara Shuzo) and overlaid with mineral oil. The resultant sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 - 55 °C for 1 min; and at 72 °C for 1 - 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). The reaction mixture was then subjected to phenol/chloroform extraction and ethanol precipitation to obtain amplified DNA fragments. The ethanol precipitation was generally carried out by adding 2.5 volumes of ethanol and either of about 1/10 volume of 3 M sodium acetate or an equal volume

of 4 M ammonium acetate to the aqueous solution, mixing, centrifuging at 15,000 rpm for 15 min using a rotor of about 5 cm in diamet r under cooling at 4 °C to pellet the precipitates, and drying the p llet. Throughout the specification, the procedure "etanol precipitation" meanes the above-mentioned procedures. In the same manner as the above, various DNA fragments were obtained using different pair of primers in PCR.

[3] Synthesis of cDNA Using Antisense Primer

To 2 μ I of RNA sample solution prepared in above [1] was added 1 μ I of 15 pmol/ μ I anti-sense primer (synthesized primer AS1, AS2 or AS3), 2 μ I of 10 x RT buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl), 4 μ I of 25 mM MgCl₂, 8 μ I of 2.5 mM 4dNTPs, 1 μ I of water and the mixture incubated at 65 °C for 5 min then at room temperature for 5 min. To the mixture was added 1 μ I of reverse transcriptase (25 U, Life Science), 1 μ I of ribonuclease inhibitor (100 U/ μ I, Takara Shuzo) and the mixture incubated at 37 °C for 20 min, at 42 °C for 30 min, and finally at 95 °C for 2 min, which was followed by an immediate cooling to 0 °C (synthesis of cDNA).

Amplification of DNA containing specific sequences was conducted by PCR (Saiki et al., Nature 324: 126 (1986)). Thus, 100 μl mixture containing 10 μl of cDNA solution, 10 μl of 10 x PCR buffer (100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 μl of 2.5 mM 4 dNTPs, 2 μl of 15 pmol/μl synthetic DNA primer (the same primer as used in the synthesis of cDNA), 3 μl of 15 pmol/μl synthetic DNA primer (a counterpart of paired primers) and water was incubated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, it was mixed with 0.5 μl of Taq DNA polymerase (7 U/μl, AmpliTaq TM Takara Shuzo) and overlaid with mineral oil. The resultant sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 - 55 °C for 1 min; and at 72 °C for 1 - 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). Finally, the reaction mixture was incubated at 72 °C for 7 min, which was followed by phenol/chloroform xtraction and ethanol precipitation to obtain different amplified DNA fragments derived from either of above-mentioned pairs of primers.

Example 3

30

Cloning and Sequencing of Amplified DNA Fragments

Dried DNA fragment (at least 1 pmole) obtained in the above Example 2, [2] or [3] was blunt-ended with T4 DNA polymerase (Toyobo) and 5'-end phosphorylated with polynucleotide kinase (Toyobo) and ligated into Smal site of multi-cloning sites of 5 ng to 10 ng of pUC19 cloning vector. The cloning vector had been previously treated as follows: digestion with a restriction enzyme Smal (Toyobo), phenol/chloroform extraction, ethanol precipitation, 5'-end dephosphorylation with alkaline phosphatase (Behringer-Manheim), phenol/chloroform extraction, and ethanol-precipitation. The ligated DNA was used to transform into a competent E.coli JM 109 or DH5 cells (Toyobo). The transformation was carried out according to the protocol of the manufacture's instruction (COMPETENT HIGH, Toyobo). Plasmid clones were recovered from transformed cells conventionally. At least 20 transformants were obtained using pUC19 cloning vectors containing either of DNA fragments obtained in the above Example 2, [2] and [3] using each pair of primers.

The determination of base sequence of DNA fragment was conducted by Fluorescent DNA Sequencer (GENESIS 2000, Dupont) using, as sequence primer, the following synthetic primers:

5' d(GTAAAACGACGGCCAGT)3' (SEQ ID NO 143) and

5'd(CAGGAAACAGCTATGAC)3' (SEQ ID NO 144) for the + and - strands of DNA fragment to be sequenced.

Base sequences of clones is given in SEQ ID NO 1 to 4 and 9 to 12. Base sequences of SEQ ID NO 1, 2, 3, 4, 9, 10, 11 and 12 correspond to that of + strand of clones N1-1, N2-1, N3-1, N10-1, N1-2, S1-1, S1-2 and S1-3 of transformants, respectively. These clones are double stranded DNA which were prepared in the same manner as those described in Examples 2 and 3 using 4 kinds of pairs of primers shown in Example 2, [2]. Plasmid used for sequencing the clones were designated as pUCN1-1, pUCN2-1, pUCN3-1, pUCN10-1, pUCN1-2, pUCS1-1, pUCS1-2 and pUCS1-3, respectively. Each plasmid contained one DNA molecule corresponding to each DNA fragment.

These base sequences represents bas s quenc s of clones obtained by cloning the cDNA synthesized from RNA isolated from serum of pati nt(s) suffering from HC. Therefore, these sequences are specific for clones originated from serum of HCV-infected patients but can not be found or obtained from serum of healthy subjects. Thus, cDNA prepared from RNA (if there are any) obtained from a healthy

subject under more strict conditions, for instance, by increasing (3 or 4 folds) the reaction cycles of PCR in Example 2, [2] and [3], by repeating them 60 - 100 times, did not show any homology in base s quence with those shown in SEQ ID NO 1 to 4. Consequently, base sequences of clones N1-1, N2-1, N3-1, N10-1, N1-2, S1-1, S1-2 and S1-3 are specific for those obtained from serum of patients suffering from HC.

As the next step, the resultant DNA fragment was modified so that a polypeptide encoded by a open reading frame should be expressed in a host cell transformed by the modified DNA, and the resultant product was then evaluated as to the ability to react, as a antigenic polypeptide of HCV, with HCAb in serum of HC patients.

o Example 4

15

Preparation of Clone N1N3N10 or N3N10

[1] Preparation of Clone N3N10

One µI of each DNA fragments (about 200 - 300 ng) from clones N3-1 and N10-1 was added into a reaction mixture containing 10 µl of 10 x PCR buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 μl of 2.5 mM 4 dNTPs, 5 μl each of 20 pmol/μl synthetic primers S2 and AS3, and 76.5 µl of water. After an intimate mixing, the mixture was heated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq™ Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 °C for 1 min; and at 72 °C for 2 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by an incubation at 97 °C for 2 min. The mixture was immediately cooled to 0 °C, kept at 0 °C for 2 min, mixed with 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq™ Takara Shuzo). The sample was then treated in the same manner as the above by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 50 °C for 1 min; and at 72 °C for 2 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The amplified DNA samples were fractionated on agarose gel electrophoresis and a gel containing a desired fragment having an expected length was removed (Molecular Cloning (1982) Cold Spring Harbor) to isolate the DNA fragment therefrom conventionally. The resultant DNA fragment was then modified as described in Example 3 and ligated into Smal site of multi-cloning sites of pUC19, cloned and screened as described in Example 3 to obtain plasmid pUCN3N10. The resultant cDNA derived from serum of HC patient was referred to as clone N3N10 whose base sequence is given in SEQ ID NO 5.

[2] Preparation of Clone N1N3N10

Two overlapping clones N1-1 and N3N10 were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme BssH11, clone N1-1 is cleaved at the 3' site of a nucleotide No. 455 G and clone N3N10 at the 3' site of a nucleotide No.159 G. The ligation of two clones N1-1 and N3N10 was accomplished on the basis of an assumption that plasmids pUCN1 and pUCN3N10 contain each clone in the same orientation. Thus, plasmid pUCN1 was digested with HindIII and BssHII to yield a 492 bp DNA fragment comprising a HindIII-Smal DNA fragment of plasmid pUC19 attached to the 5' end of the No. 455 bp nucleotide of clone N1-1 derived from serum of HC patient, which fragment was then exchanged with 159 bp HindIII - BssHII fragment of Plasmid pUCN3N10, cloned and screened to obtain a plasmid pUCN1N3N10. The plasmid pUCN1N3N10 contained the desired clone N1N3N10 comprising clones N1-1, N3-1 and N10-1 ligated without overlapping. The base sequence of clone N1N3N10 is shown in SEQ ID NO 6.

Example 5

Modification of DNA for the Expression of HCV Polypeptide Encoded by Clones N3-1 or N3N10

[1] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clone N3-1 in E.coli

Clone N3-1 contains a DNA fragment capable of encoding a structural protein of HCV which begins at nucleotide No. 22 (A). The DNA can be expressed utilizing ATG codon at nucleotides Nos. 22 to 24. The modification of DNA was carried out using PCR. The following synthetic oligonucleotide primers were used.

25

55

50

- 5' primer:
- 5' GCAAGCTTATGAGCACAAATCCAAAACCCCAAAGA 3' (SEQ ID NO 145)
- 3' primer:
- 5' GCGAATTCAGATCTTCACCTACGCCGGGGGTCCGTGGG 3' (SEQ ID NO 146)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as described in the above according to Saiki's method in a total volume of 100 μl containing 100 ng of plasmid pUCN3, as a template, and 2 μl each of 3' and 5' primers. The reaction mixture was heated at 95 °C for 5 min and quenched at 0 °C. One minute later, to the mixture was added 0.5 μl of Taq DNA polymerase (7 U/ml, AmpliTaq TM Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 25 cycles of treatments which comprises: at 95 °C for 1 minute; at 60 °C for 1 min; and at 72 °C for 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). The resultant reaction solution was extracted with phenol/chloroform, and precipitated with ethanol conventionally. The amplified DNA samples were digested with HindIII and EcoRI, and fractionated on acrylamide gel electrophoresis and extracted (Molecular Cloning, Cold Spring Harbor (1982)).

The resultant DNA fragment was then ligated into HindIII and EcoRI sites of a cloning vector pUC19, cloned and screened to obtain plasmid pUCHN3. The resultant plasmid was sequenced and shown in SEQ ID NO 7. The sequence shows that it contains, at the 5'-terminus, a HindIII site followed by ATG initiation codon, and at the 3'-terminus, a termination codon TGA, BgIII and EcoRI restriction sites, from 5' to 3'.

[2] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clone N3N10 in E.coli

Clone N3N10 contains a DNA fragment capable of encoding structural protein of HCV which begins at nucleotide No. 22 (A). The DNA can be expressed utilizing ATG codon at nucleotides Nos. 22 to 24. The modification of DNA was carried out using PCR. The following synthetic oligonucleotide primers were used.

- 5' primer:
- 5' GCAAGCTTATGAGCACAAATCCAAAACCCCAAAGA 3' (SEQ ID NO 145)
- 3' primer:

30

5' GCGAATTCAGATCTTCAGATTCTCTGAGACGGCCCTCGT 3' (SEQ ID NO 147)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as the above [1] except that the above two primers and plasmid pUCN3N10, as a template, were used and PCR was conducted by repeating 10 cycles of treatments which comprises: at 95 °Cfor 1 minute; at 50 °C for 1 min; and at 72 °C for 5 min, and then 20 cycles of treatments which comprises: at 95 °Cfor 1 minute; at 65 °C for 1 min; and at 72 °C for 5 min.

The amplified DNA sample was digested with HindIII and EcoRI, and fractionated on acrylamide gel electrophoresis and extracted the gel containing a DNA fragment of desired length (Molecular Cloning, Cold Spring Harbor (1982)). The resultant DNA fragments were then ligated into HindIII and EcoRI sites of cloning vector pUC19, cloned and screened conventionally to obtain plasmid pUCHN3N10. The plasmid pUCHN3N10 was then sequenced.

Thus obtained clone HN3N10 contains, at the 5'-terminus, a HindIII site followed by ATG initiation codon, and at the 3'-terminus, a termination codon TGA, BgIII and EcoRI restriction sites, from 5' to 3'.

For the removal or BamHI site from the clone HN3N10, a nucleotide sequence: 5'GGATCC3' was converted to 5'GGATAC3' by PCR using the following synthetic DNA fragments as primers.
5' primer:

- 5' GCTACTCCGGATACCAC 3' (SEQ ID NO 148)
- 5 3' primer:
 - 5' GTAAAACGACGGCCAGT 3' (SEQ ID NO 143)

The synthetic DNA was adjusted to 20 pmol/ml before use.

The nucleotide "G" at the 5'-terminus of 5' primer corresponds to the No.1016 G of the base sequence of clone N3N10. The 3' primer is derived from plasmid pUC19 and the same as one of primers used for sequencing in Example 3. The PCR was conducted by repeating 25 cycles of treatments which comprises: at 95 °C for 1 minute; at 55 °C for 1 min; and at 72 °C for 1 min. For the reaction, 3 µl of each primer and 100 ng of plasmid pUCHN3N10, as a template DNA, were used. The reaction mixture was then subjected to phenol/chloroform extraction and ethanol precipitation as conventionally. The amplified DNA sample was digested with Mrol, BgIII, and BamHI, fractionated on acrylamide gel electrophoresis, and extracted the gel containing a desired 226 bp DNA fragment (Molecular Cloning, Cold Spring Harbor (1982)). The resultant DNA fragments were then ligated into Mrol and BgIII sites of plasmid pUCHN3N10, cloned and screened by conventional method to obtain plasmid pUCHN3N10AB. The resultant plasmid pUCHN3N10AB was then sequenced and base sequence of clone HN3N10AB is shown in SEQ ID NO 8.

[3] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clone N3N10 in Insect Cells

Clone N3N10 appears to contain entire viral protein-encoding genes including those encoding core, nvelope (M-gp35) proteins. The region beginning at nucl otide No. 22 (A) which ncodes structural protein was expressed in insect cells utilizing ATG codon at nucleotides Nos. 22 to 24. When insect cells were transfected with the DNA and cultivated, core andenvelope (M-gp35) proteins were expressed in the fused form as a precursor polypeptide, which was then processed to separate core and envelope (M-gp35). At least the latter envelope (M-gp35) was then glycosylated incompletely and accumulated intracellurarly. The modification of DNA of clone N3N10 for the construction of expression vector was carried out by PCR using following synthetic oligonucleotide primers.

5' primers:

MS106: 5' GCGTCGACGCTAGCATGAGCACAAATCCAAAACCC 3' (SEQ ID NO 149)

MS107: 5' GCGTCGACGCTAGCAGGTCTCGTAGACCGTGCATC 3' (SEQ ID NO 150)

3' primer:

MS108: 5' GCGAATTCGCTAGCTCAGGATTCTCTGAGACGGCCCTCGA 3' (SEQ ID NO 151)

These three synthetic DNAs were separately adjusted to 20 pmol/ml before use.

The PCR was carried out using the same reaction solution and worked up in the same manner as described in the above [1] except that plasmid pUCN3N10 was used as a template plasmid, and, as 5'primer, primer MS106 or MS107 and, as 3'primer, MS108 were used. PCR was accomplished by repeating 10 times of reaction cycles consisting of: 1 min at 95 °C; 1 min at 50 °C and 5 min at 72 °C; and then 20 times of reaction cycles consisting of: 1 min at 95 °C; 1 min at 65 °C; 5 min at 72 °C. A combination of primers MS106 and MS108 gave a desired 1265 bp DNA fragment 106-108 and that of primers MS107 and MS108 gave a desired 1286 bp DNA fragment 107-108.

These DNA fragments were digested with Nhel, fractionated on acrylamide gel electrophoresis and extracted by convenional means (Molecular Cloning, Cold Spring Harbor (1982)) to obtain DNA fragments of desired length. Each of the resultant DNA fragments was then ligated into Nhel restriction site of a transfer vector pBlueBac (Invitrogen), cloned and screened by the usual method to yield plasmids pBlueN3N10-1 and pBlueN3N10-2, which are derived from DNA fragments 106-108 and 107-108, respectively.

Plasmids pBlueN3N10-1 and pBlueN3N10-2 were digested with Nhel or BamHl completely to confirm that each plasmid contains only one DNA fragment, either of 106-108 or 107-108 inserted at Nhel site. Furthermore, taking account of the instruction provided by the manufacture (Invitrogen), the expression unit of these plasmid contain a gene encoding HCV structural polypeptide (core and envelope) oriented forward and ligated to the Nhel cloning site down stream from a polyhedrin promoter.

35 Example 6

Expression of HCV Polypeptides Encoded by Clones HN3, HN3N10∆B

[1] Expression of Polypeptide Encoded by Clone HN3 in E.coli

Clone HN3 encodes a part of polypeptide encoded by cDNA originated from serum of HC patient. The polypeptide encoded by clone HN3 was expressed directly in E.coli, as it is, by subcloning said clone into an expression vector pCZ44 (Japanese Patent Publication No. 124387/1989).

Clone HN3 was digested thoroughly with restriction enzymes HindIII and BgIII, extracted with phenol/chloroform, precipitated with ethanol, separated on acrylamide gel electrophoresis. From the gel was extracted a DNA fragment having cohesive HindIII- and BgIII-restricted ends (Molecular Cloning, Cold Spring Harbor, 1982). The expression vector pCZ44 was digested with HindIII and BgIII. The larger DNA fragment containing a region functional for the replication in E.coli was separated, treated in the same manner, ligated to the HindIII-BgIII fragment of clone HN3 so as to have only one insertion, and cloned by conventional method to yield plasmid pCZCORE.

Alternatively, an expression vector was constructed using an expression vector pGEX-2T (Pharmacia) for the expression of a fused protein of a desired polypeptide and β-glutathione-S-transferase (GST). The construction was carried out substantial in accordance with the protocol taught by the manufacture (Pharmacia). Thus, the expression vector pGEX-2T was digested with BamHI. The linearized vector was ligated with a HindIII linker to obtain a DNA fragment having EcoRI and HindIII restriction sites at the 3'- and 5'-termini. The fragment was ligated to HindIII-EcoRI fragment of HN3 such that every reading frame of codon is consistent with an amino acid of clone N3-1 to yield an expression vector pGEXCORE.

E.coli K12 strains (e.g., JM109, KS476) or those derived from B strains transformed with plasmid

pCZCORE was grown in L-Broth at 37 °C overnight (Molecular Cloning, Cold Spring Harbor, 1982). The cultured broth was diluted 50-folds by inoculating it into a freshly prepared L-Broth and the cultivation continued with shaking at 30 °C for 2 hr. At this time, IPTG (isopropyl-β-D-galactopyranoside) was added to the culture to a final concentration of 1 or 2 mM in order to induce the expression of DNA encoding HCV-originated CORE-N3 polypeptide by single-clone-derived transformants (E.coli cells transformed solely by plasmid pCZCORE derived from clone HN3). Base sequence and deduced amino acid sequence of clone HN3 is shown in SEQ ID NO 7.

As mentioned in the above, plasmid pGEXCORE can be used to obtain transformants capable of expressing a fused protein include desired polypeptide and GST. The plasmid encodes a fused protein GST-CORE comprising GST, which has a thrombin-cleaving site at its C-terminus, and a polypeptide derived from a clone HN3, the same polypeptide as that encoded by plasmid pCZCORE. The transformants containing pGEXCORE were grown in the presence of IPTG using the same protocol as that used for the expression of CORE-N3 polypeptide of HCV from transformants harboring pCZCORE to produce the fused polypeptide GST-CORE.

[2] Expression of Polypeptides Encoded by Clone HN3N10∆B

15

Clone HN3N10ΔB encoding a part of polypeptide encoded by cDNA originated from serum of HC patient was expressed in E.coli to give polypeptide CME-N3N10ΔB in the same manner as the above [1]. The cDNA used was that contained in clone HN3N10ΔB obtained from serum of HC patient, which clone had been previously isolated and sequenced as described in Examples 3, Example 4 [1], and Example 5 [2]. The expression plasmid pCZCMEΔB was constructed by subcloniong a DNA fragment isolated from plasmid pUCHN3N10ΔB by ligating its HindIII and BgIII cohesive ends to HindIII and BgIII sites of plasmid pCZ44 such that only one DNA fragment should be inserted in an appropriate orientation by the same method used for the preparation of plasmid pCZCORE. Plasmid pCZCMEΔB was then subjected to the sequencing and restriction enzyme mapping to confirm that an expression unit of plasmid pCZCMEΔB was reconstructed properly.

The cultivation of transformants was carried out in the presence of IPTG in order to induce the expression of HCV-originated CME-N3N10ΔB polypeptide by single-clone-derived transformants (E.coli JM 109 cells transformed solely by plasmid pCZCMEΔB derived from clone HN3N10ΔB, a variant of clone N3N10). Base sequence and deduced amino acid sequence of cDNA obtained from serum of HC patient contained in clone HN3N10ΔB is shown in SEQ ID NO 8. The amino acid sequences deduced from base sequences of a clone HN3N10ΔB and its original clone N3N10 were exactly the same.

In the same manner as the above [1], plasmid pGEXCMEAB was constructed, transformed into host cells. The transformants, when grown under a same condition for transformants harboring plasmid pCZCMEAB inducing by IPTG, expressed a fused protein GST-CME-N3N10AB.

[3] Expression of Polypeptide Encoded by Clone N3N10 in Insect Cells

The expression of structural polypeptide (core, envelope (M-gp35) of HCV encoded by plasmid pBlueN3N10-1 prepared in Example 5 [3] was conducted substantial in accordance with a known expression manual for baculovirus (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4, hereinafter, referred to as Maxbac, Invitrogen).

Plasmids pBlueN3N10-1 and pBlueN3N10-2, plasmids prepared by inserting DNA fragment containing HCV structural gene at the Nhel site of a transfer vector pBlueBac (Maxbac, pp.37), were recovered from E.coli host cells transformed thereby, and purified according to the method of Maniatis et al.(Molecular Cloning, Cold Spring Harbor Laboratory, pp.86 - 96 (1982)). Thus, a large amount of HCV structural genecontaining transfer plasmid DNA was obtained. Sf9 cells were co-transfected with 2 µg of either of plasmids pBlueN3N10-1 or pBlueN3N10-2 and 1 µg of AcNPV viral DNA (Maxbac, pp.27). Sf9 cells were grown in TMN-FH medium (Invitrogen) containing 10% FCS (fetal calf serum) in a 6 cm dish (60 x 15 mm, FALCON^R; Nippon Becton Dickinson Co., Ltd.) until a cell density reached to about 2 x 10⁶/plate. The TMN-F medium was removed and a 0.75 ml Grace medium (Gibco) containing 10% FCS was added thereto. To the DNA mixture described in th above was added 0.75 ml of transfection buffer (attached to the kit) was thoroughly mixed by vortex and gradually added dropwise onto the Grace medium. After the culture being allowed to stand for 4 hr at 27 °C, Grace m dium was replaced with 3 ml of TMN-FH medium containing 10% FCS and the dish incubated at 27 °C for 6 days. Three days from the incubation, there observed a few multinucleate cells and on sixth day, almost all the cells were multinuclear. The supernatant was taken into a centrifuging tube and centrifuged at 1,000 rpm, 10 min to obtain the supernatant as a cotransfected viral

solution.

The cotransfected viral solution contains about 10⁸ viruses/ml and 0.5% of which were recombinant viruses. The isolation of recombinant virus was carried out by a plaque isolation method described below.

Thus, cells were adsorbed onto a 6 cm dish by se ding 1.5×10^6 cells on medium and removing the medium completely. To the dish was added $100 \, \mu l$ of a diluted viral solution (10^{-4} and 10^{-5} folds), separately and incubated at room temperature for 1 hr while slanting the 6 cm dish every 15 min to spread the virus extensively. X-gal medium containing agarose was prepared by adding 5-bromo-4-chloro-3-indolyl- β -D-galactoside to a final concentration of 150 $\mu g/l$ (Maxbac, pp. 16-17) to a warm medium which had been prepared by autoclaving 2.5% baculovirus agarose (Invitrogen) at 105 °C for 10 min, mixing with TMN-FH medium containing 10% FCS preheated at 46 °C at the mixing ratio of 1:3, and keeping the temperature at 46 °C.

After the completion of infection, virus solution was aspirated thoroughly from the 6 cm dish and 4 ml of the warm X-gal medium containing agarose (previously prepared) was gently added to every 6 cm dish not to peel off cells. The dish kept open by slightly sliding a lid until the agarose solidified and dried, and thereafter the dish covered, turned upside down, and incubated at 27 °C for 6 days. The plaques were observed under a phase difference microscope to find blue plaques which do not form multinucleate cells. Agarose containing blue recombinant plaques were removed with a Pasteur pipet and suspended into 1 ml of TMN-FH medium by pipetting many times. The above process which comprised: infection, 6-day incubation, and isolation of virus containing transfer plasmid DNA is called the "plaque method". The plaque method was repeated using 100 µl of viral suspension. After repeating said process three times, there obtained a recombinant virus having a gene encoding structural protein derived from HCV free from contamination with that of wild-type strain.

A viral solution of the primary recombinant virus was prepared by aspirating plaques with a Pasteur pipet, and mixing thoroughly with 1 ml of TMN-FH medium. Because the primary viral solution was low in virus density for infection, it required further treatments for concentration. Thus, $100~\mu l$ of viral solution was adsorbed onto Sf9 cells grown in 6 cm dish to a semi-confluent, and 4 ml of TMN-FH medium was added thereto and incubated three days. The culture supernatant was recovered to yield a recombinant viral solution for infection.

For the production of HCV structural protein, a suspension of Sf9 cells in TMN-FH medium containing 10% FCS (5 x 10⁶ cells/10 ml medium) was added into a 9 cm dish and kept 1 hr for adsorption. After the removal of medium, 250 μ l of recombinant viral solution was added to the 9 cm dish and spread extensively. To the dish was added 10 ml TMN-FH medium containing 10% FCS and incubated at 27 °C for 4 days. The cells expressing recombinant glycoprotein of HCV were harvested by scraping up and suspended into 1,000 ml of phosphate buffered saline.

Thus, HCV structural gene was expressed in Sf9 cells transfected with said virus. The transformants transformed with plasmids pBlueN3N10-1 and pBlueN3N10-2 expressed the same HCV polypeptide.

Example 7

35

50

Identification of Expression Products as HCAg

The expression products obtained in Example 6, which are CORE-N3 and CME-N3N10 Δ B polypeptides, and HCV polypeptide encoded by clone N3N10 expressed in insect cells, were immunologically reactive with antiserum obtained from HC patients, demonstrating that these expression products are HC associated antigens.

Identification of these expression products as HCAg were carried out by Western blot as follows. E. colicells transformed with either of plsmids pCZCORE and pCZCMEΔB encoding CORE-N3 and CME-N3N10ΔB polypeptides, respectively were grown under the presence of IPTG for 3 hr or a overnight in the same manner as described in Example 6.

Recombinant strains were harvested by centrifuging 1,000 μ I of the cultured broth at 6,500 rpm, 10 min. The pellet was dissolved into a sample solution (50 mM Tris-HCl, pH6.8 containing 2% SDS, 5% mercaptoethanol, 10% glycerin, and 0.005% bromophenol blue) for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml. Sf9 cells infected with viruses which had been treat d more than 3 tim s by plaque method were collected by scraping up and suspended into 1,000 ml of phosphate-buffered saline physiological saline (PBS) and 100 μ I of the suspension was centrifuged at 6,500 rpm, 10 min to pellet th cells. The pellet was dissolved into a sample solution for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml.

The sample solutions were then boiled at 100 °C for 10 min. Ten µI of the boiled solution was loaded

onto 0.1% SDS-15% polyacrylamide gel (70 x 85 x 1 mm) together with a marker protein LMW Kit E (low-molecular weight marker prot in, Pharmacia). Electrophoresis was carried out at a constant current of 30 mA for 45 min in Tris buffer (25 mM Tris, pH 8.3, 192 mM glycine, 0.1% SDS) as electrode buffer. Thereafter, DNA was transferr d electophoretically to a nitrocellulose filter by superposing the gel onto a filter BA-83 (S & S), impressing a constant current of 120 mA for about 20 min between gel (cathode) and the filter (anode) as conventionally.

The transcribed filter was cut to remove a part containing a marker protein (referred to as marker filter) and that containing the sample (referred to as sample filter). The former was stained with 0.1% (w/v) amideblack 10B and the latter immersed into 0.01 M PBS (pH 7.4) containing 5% (w/v) bovine serum albumin (BSA). Serum from a HC patient was diluted 50 times with 0.01 M PBS (pH 7.4) containing 5% (w/v) BSA. To the sample filter was added 10 µl of diluted serum and the filter allowed to stand for 2 hr at room temperature. Thereafter, the filter was washed with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3).

The sample filter was then reacted with 10 ml of horseradish peroxidase conjugated anti human IgG (Gappel) at 37 °C for 1 hr and washed with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3). The filter was then immersed into peroxidase-color-producing solution (60 mg 4-chloro-1-naphthol, 20 ml methanol, 80 ml PBS, and 20 µl aqueous hydrogen peroxide). The colored filter was washed with distilled water and compared with the marker filter, demonstrating that polypeptides CORE-N3 and CME-N3N10ΔB contain only one colored protein having a reasonable molecular weight as an expression product of cDNAs originated from serum of HC patients and contained in plasmid pCZCORE and pCZCMEΔB, respectively.

Cells transformed with pBlueN3N10-1 or plasmid pBlueN3N10-2, both of which encode polypeptide encoded by clone N3N10, expressed HCV polypeptides showing the same pattern on the detection. A protein of molecular weight of about 22 kD was expressed which corresponds to calculated molecular weight of an expression product from core-encoding gene contained in clone N3N10. Thus, said core-encoding gene, when expressed, gives a protein of calculated molecular weight of about 22 kD (without modification). As the result, the expressed product was identified as hepatitis C associated antigenic polypeptide presumably derived from HCV core protein.

Example 8

30

Comparison of Clones Obtained in Example 2 [2] and [3]

Three clones corresponding to SEQ ID NO 1 were separately cloned using serum from a HC patient according to the method described in Example 2 [2] (using random primers) and sequenced. On the other hand, three clones corresponding to SEQ ID NO 1 were separately cloned using serum from the same HC patient according to the method described in Example 2 [3] (using antisense primers) and sequenced.

Clones obtained using random primers had the same base sequence as that shown by SEQ ID NO 1, whereas the synthetic primers S1 and AS1 were used, two of three clones obtained independently had the base sequence of SEQ ID NO 1, and one clone had a base sequence which differed from that of SEQ ID NO 1 as to three nucleotides. Thus, at No. 345, A was changed to C, No.322 A changed to T, and No. 95 A changed to C. These differences indicate that a patient is infected at least 2 kinds of viruses.

The above facts demonstrate that there are no substantial difference between clones obtained by methods in Example 2 [2] and those obtained in Example 2 [3].

45 Example 9

50

55

Synthesis of cDNA

[1] Preparation of RNA Sample Solution

RNA sample solution was prepared by resolving the dried nucleic acid obtained in Example 1 in 30 μ l of water containing 10 μ l of ribonuclease inhibitor (100 U/ μ l, Takara Shuzo, Japan).

[2] Synthesis of cDNA Using Antisense Primer

To 2 μI of RNA sample solution prepared in above [1] was added 1 μI of 15 pmol/μI anti-sense primer (synthesized primer MS122, MS157 or MS148), 2 μI of 10 x RT buffer (100 mM Tris-HCI (pH8.3), 500 mM KCI), 4 μI of 25 mM MgCl₂, 8 μI of 2.5 mM 4dNTPs, 1 μI of water and the mixture incubated at 65 °C for 5

min then at room temperature for 5 min. To the mixture was added 1 μ I of reverse transcriptase (25 U, Life Science), 1 μ I of ribonuclease inhibitor (100 U/ μ I, Takara Shuzo) and the mixture incubated at 37 °C for 20 min, at 42 °C for 30 min, and finally at 95 °C for 2 min, which was followed by an immediate cooling to 0 °C (synthesis of cDNA).

Amplification of DNA containing specific sequences was conducted by PCR (Saiki et al., Nature 324: 126 (1986)). Thus, 100 µl mixture containing ten µl of cDNA solution, 10 µl of 10 x PCR buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 µl of 2.5 mM 4 dNTPs, 2 µl of 150 pmol/µl synthetic DNA primer (the same primer as used in the synthesis of cDNA), 3 µl of 15 pmol/µl synthetic DNA primer (a counterpart of pair of primers, i.e.,MS122-MS123, MS157-MS156, or MS148-MS146) and water was incubated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, it was mixed with 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq ™ Takara Shuzo) and overlaid with mineral oil. The resultant sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 - 55 °C for 1 min; and at 72 °C for 7 min in DNA Thermal Cycler (Parkin Elmer Cetus). Finally, the reaction mixture was incubated at 72 °C for 7 min, which was followed by phenol/chloroform extraction and ethanol precipitation to obtain different amplified DNA fragments derived from either of above-mentioned pairs of primers.

Example 10

Cloning and Sequencing of Amplified DNA Fragments

Dried DNA fragment (at least 1 pmole) obtained in the above Example 9, [2] was blunt-ended with T4 DNA polymerase (Toyobo) and 5'-end phosphorylated with polynucleotide kinase (Toyobo) and ligated into Smal site of multi-cloning sites of 5 ng to 10 ng of pUC19 cloning vector. The cloning vector had been previously treated as follows: digestion with a restriction enzyme Smal (Toyobo), phenol/chloroform extraction, ethanol precipitation, 5'-end dephosphorylation with alkaline phosphatase (Behringer-Manheim) (Molecular Cloning (1982) Cold Spring Harbor), phenol/chloroform extraction, and ethanol-precipitation. The ligated DNA was used to transform a competent E.coli JM 109 or DH5 cells (Toyobo). The transformation was carried out according to the protocol of the manufacture's instruction (COMPETENT HIGH, Toyobo). Plasmid clones were recovered from transformed cells conventionally. At least 20 transformants were obtained using pUC19 cloning vectors containing either of DNA fragments obtained using either of pairs of primers in the same manner as that described in Example 9, [2].

Plasmid DNA was isolated from corresponding transformant by an usual method and sequenced. The determination of base sequence was conducted by means of Fluorescent DNA Sequencer (GENESIS 2000, Dupont) using, as sequence primer, the following synthetic primers:

5' d(GTAAAACGACGGCCAGT)3' (SEQ ID NO 143) and

5'd(CAGGAAACAGCTATGAC)3' (SEQ ID NO 144) for the + and - strands of DNA fragment to be sequenced.

Base sequences of DNA fragments are given in SEQ ID NO 13 to 27, which show the base sequences of + strand of HCV genes inserted into each plasmid used for the transformation. These clones are double stranded DNA. Plasmids used for the sequencing of clones N19-1, N19-2 and N19-3 were designated as plasmids pUCN19-1, pUCN19-2 and pUCN19-3, respectively. Each plasmid contained one DNA molecule corresponding to each DNA fragment. In the same manner, a plasmid which contains a single clone and is used for the sequencing of the same is designated by adding a prefix "pUC" to the name of the clone.

These base sequences represents base sequences of clones obtained by cloning the cDNA synthesized from RNA isolated from serum of patient(s) suffering from HC. Therefore, these sequences are specific for clones originated from serum of HCV-infected patients but can not be found or obtained from serum of healthy subjects. Thus, cDNA prepared from RNA (if there are any) obtained from a healthy subject under more strict conditions, for instance, by increasing (3 or 4 folds) the reaction cycles of PCR in Example 9 [2] and [3], by repeating them 60 - 100 times, did not show any homology in base sequence with those shown in SEQ ID NO 13 to 27. Consequently, base sequences of clones shown in SEQ ID NO 13 to 27 are specific for those obtained from serum of HC patient.

The base sequences of DNA fragments were compared with a known base sequence of HCV gene. As can be seen from the fact that three clones N19-1, N19-2 and N-193 were obtained from serum of one HC patient in Example 9 [2] using primers MS122 and MS123, there must be more than one virus in a patient.

Example 11

Preparation of Clones N27MX24A-1 and N27MX24B-1

[1] Preparation of Clones N19MX24A-1 and N19MX24B-1

One µI (about 0.5 to 1 µg/µI) of each DNA fragment from clones N19-1 and MX24-4 was added into a reaction mixture containing 10 μl of 10 x PCR buffer (100 mM Tris-HCI (pH8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 µl of 2.5 mM 4 dNTPs, 5 µl each of 20 pmol/µl synthetic primers S2 and AS3, and 76.5 µl of water. After an intimate mixing, the mixture was heated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq™ Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 °C for 1 min; and at 72 °C for 2 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by an incubation at 97 °C for 2 min. The mixture was immediately cooled to 0 °C, kept at 0 °C for 2 min, mixed with 0.5 μl of Taq DNA polymerase (7 U/μl, AmpliTaq™ Takara Shuzo). The sample was then treated in the same manner as the above by repeating 25 times of reaction cycle, which comprises the following treatments: at 95°C for 1 min; at 50 °C for 1 min; and at 72 °C for 2 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The amplified DNA samples were fractionated on agarose gel electrophoresis and a gel containing a fragment having a desired length was removed (Molecular Cloning (1982) Cold Spring Harbor) to isolate the DNA fragment therefrom conventionally. The resultant DNA fragment was then modified as described in Example 10 and ligated into Smal site of multi-cloning sites of pUC19, cloned and screened as described in Example 10 to obtain plasmids pUCN19MX24A-1 and pUCN19MX24B-1. The resultant cDNAs derived from serum of HC patient were referred to as clones N19MX24A-1 and N19MX24B-1, of which base sequences are given in SEQ ID NO 29 and 30.

[2] Preparation of Clone N27N19-1

25

30

40

Two overlapping clones N27-3 and N19-1 were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme Mlul, clone N27-3 is cleaved at the 3' site of a nucleotide No. 330 (A) and clone N19-1 at the 3' site of a nucleotide No.51 (A). The ligation of clones N27-3 and N19-1 was accomplished on the basis of an assumption that plasmids pUCN27-3 and pUCN19-1 contain each DNA fragment in the same orientation. Thus, plasmid pUCN27-3 was digested with HindIII and Mlul to isolate a DNA fragment containing 5' region of clone N27-3 which comprises a HindIII-Smal DNA fragment of plasmid pUC19 attached to the 5' end of the clone N27-3, a cDNA derived from serum of HC patient. The DNA fragment was then exchanged with a HindIII-Mlul fragment of clone N19-1 containing 3' region of said clone, cloned and screened to obtain a plasmid pUCN27N19-1. The plasmid pUCN27N19-1 contained the desired clone N27N19-1 comprising clones N27-3 and N19-1 ligated without overlapping. The base sequence of clone N27N19-1 is shown in SEQ ID NO 28.

[3] Preparation of Clones N27MX24A-1 and N27MX24B-1

Overlapping clones N27-3 and either of clones N19MX24A-1 and N19MX24B-1 were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme Mlul, clone N27-3 is cleaved at the 3' site of a nucleotide No. 330 (A) and clones N19MX24A-1 and N19MX24B-1 at the 3' site of a nucleotide No.71 (A). The ligation of clones was accomplished on the basis of an assumption that plasmids pUCN27-3, pUCN19MX24A-1 and pUCN19MX24B-1 contain each DNA fragment in the same orientation. Thus, plasmid pUCN27-3 was digested with HindIII and Mlul to isolate a 363 bp DNA fragment which comprises a HindIII-Smal DNA fragment of plasmid pUC19 attached to the 5' end of the clone N27-3, a cDNA derived from serum of HC patient. The DNA fragment was then exchanged with a 363 bp DNA fragment of clone N19MX24A-1 or N19MX24B-1 which were excised from plasmids pUCN19MX24A-1 and pUCN19MX24B-1 with HindIII and Mlul restriction enzymes, followed by cloning and screening. The resultant plasmids pUCN27MX24A-1 and pUCN27MX24B-1 contained the desired clones N27MX24A-1 and N27MX24B-1, each comprising a clone N27-3 and either of clones N19MX24A-1 and N19MX24B-1 ligated without overlapping. The base s quences of clones N27MX24A-1 and N27MX24B-1 are shown in SEQ ID NO 31 and 32, respectively.

Example 12

Modification of DNA for the Expression of HCV Polypeptide Encoded by Clones N27MX24A-1 and N27MX24B-1

[1] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clones N27MX24A-1 and N27MX24B-1 in E.coli

Clones N27MX24A-1 and N27MX24B-1 appeared to encode an open reading frame from the nucleotide No.2 (C) derived from HCV gene, which can be expressed by inserting an ATG initiation codon inframe and upperstream from said gene so that the expression of the DNA might be properly effected in host cells. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of an amino acid sequence of SEQ ID NO 31 or 32. When an expression vector containing an initiation codon for E. coli. is used, a DNA fragment from the clone is ligated to the vector such that frame of said DNA is in fonfirmity with that of the ATG codon. The modification of DNA can be carried out by 15 PCR. The modification procedures are hereinafter illustrated using clone N27MX24A-1. It will be appreciated that clone N27MX24B-1 can be modified just in the same manner.

The following synthetic oligonucleotide primers were used.

5' primer:

MS2724-1; 5' GCAAGCTTATGCGGATCCCACAAGCCGTGGTGGAT 3' (SEQ ID NO 152)

5' primer for inserting DNA fragment into a vector containing initiation codon.

MS2724-2; 5' CGGATCCCACAAGCCGTGGTGGAT 3' (SEQ ID NO 153)

3' primer:

25

45

55

MS2724-3; 5' GCGAATTCAGATCTTCATCACTCTAAGGTGGCGTCGGCGTGGG 3' (SEQ ID NO 154)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as described in the above according to Saiki's method in a total volume of 100 µl containing 100 ng of plasmid pUCN27MX24A-1, as a template, and 2 µl each of 3' and 5' primers. The reaction mixture was heated at 95 °C for 5 min and quenched at 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/ml, AmpliTaq™ Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 25 cycles of treatments 30 which comprises: at 95 °C for 1 minute; at 60 °C for 1 min; and at 72 °C for 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). The resultant reaction solution was extracted with phenol/chloroform and precipitated with ethanol conventionally. The amplified DNA samples were digested with Hindlil and EcoRI (when MS2724-2 was used as 5'primer, the DNA was blunt ended with T4 DNA polymerase and digested with EcoRI), and fractionated on acrylamide gel electrophoresis and the gel containing a DNA fragment of desired length was extracted (Molecular Cloning, Cold Spring Harbor (1982)).

The resultant DNA fragment was then ligated into HindIII (when MS2724-2 was used as 5'primer, Smal) and EcoRI sites of a cloning vector pUC19, cloned and screened to obtain plasmid pUCHN27MX24A-1 (plasmid pUCH2N27MX24A-1, when MS2724-2 was used). The resultant plasmid was sequenced. Clone CHN27MX24A-1 comprises a DNA fragment shown by a base sequence of SEQ ID No 31, 32 except that the 5' terminal C was removed and the following DNA fragment:

5' GCAAGCTTATG 3'

3' CGTTCGAATAC 5' (SEQ ID NO 155)

which comprises a HindIII restriction site followed by an initiation codon ATG, was added thereto, and 3' terminal two bases (AA) were removed from the base sequence of SEQ ID NO 31 and the following DNA fragment:

TGATGAAGATCTGAATTCGC 3'

3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

which comprises two termination codons, and EcoRI sites from 5' to 3', was added thereto.

Another clone H2N27MX24A-1 obtained using primers MS2724-2 and MS2724-3 was sequenced showing that said clone has no additional DNA fragment at the 5' terminus but, at the 3' terminus, has the same additional DNA fragment as that of the above clone HN27MX24A-1.

[2] Modification of a DNA Fragment for the Expression of HCV Polypeptide Comprising 106 Amino Acid Sequence from No. 109 to 214 of SEQ ID NO 31, 32 in E.coli

A DNA fragment encoding a polypeptide comprising 106 amino acid sequence from Nos. 109 to 214 amino acids of SEQ ID NO 31, 32 appeared to encode an open reading frame (ORF) from HCV gene, which can be expressed by inserting an ATG initiation codon in frame and upperatream from said gene. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide to the N' terminus (amino terminus) of said polypeptide. When an expression vector containing an initiation codon for E. coli. is used, a DNA fragment from the clone is ligated to the vector such that the frame of said DNA is in confirmity with that of the codon. The modification of DNA can be carried out by PCR using the following synthetic oligonucleotide primers. 5' primer:

MSHNS1-1: 5' GCAAGCTTATGTTCAACGCGTCCGGATGTCCGGA 3' (SEQ ID NO 157)

5' primer for inserting DNA fragment into a vector containing initiation codon.

MSHNS1-2: 5' TTCAACGCGTCCGGATGTCCGGA 3' (SEQ ID NO 158)

3' primer:

MSHNS1-3: 5' GCGAATTCAGATCTTCATCAACAACCGAACCAGTTGCCCTGCG 3' (SEQ ID NO 159)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as described in the above according to Saiki's method in a total volume of 100 μI containing 100 ng of plasmid pUCN27MX24A-1 (or plasmid pUCN27MX24B-1), as a template, and 2 μI each of 3' and 5' primers. The reaction mixture was heated at 95 °C for 5 min and quenched at 0 °C. One minute later, to the mixture was added 0.5 μI of Taq DNA polymerase (7 U/mI, AmpliTaqTM Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 25 cycles of treatments which comprises: at 95 °C for 1 minute; at 60 °C for 1 min; and at 72 °C for 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). The resultant reaction solution was extracted with phenol/chloroform and precipitated with ethanol conventionally. The amplified DNA samples were digested with HindIII and EcoRI (when MSHNS1-2 was used as 5'primer, the DNA was blunt ended with T4 DNA polymerase and digested with EcoRI), and fractionated on acrylamide gel electrophoresis and extracted (Molecular Cloning, Cold Spring Harbor (1982)).

The resultant DNA fragment was then ligated into HindIII (when MSHNS1-2 was used as 5'primer, Smal) and EcoRI sites of a cloning vector pUC19, cloned and screened to obtain plasmid pUCH48 (plasmid pUCH48-2, when primers MSHNS1-2 and MSHNS1-3 were used). The resultant plasmid was sequenced, demonstrating that the clone H48 has a modified base sequence of SEQ ID NO 31, 32 wherein, at the 5' site of No. 326 T, the following DNA fragment:

40

- 5' GCAAGCTTATG 3'
- 3' CGTTCGAATAC 5' (SEQ ID NO 155)

45

which fragment comprises a HindIII restriction site at 5' terminus and ATG initiation codon, followed by an initiation codon ATG, was added, and, at the 3' terminus, the following DNA fragment:

50

- 5' TGATGAAGATCTGAATTCGC 3'
- 3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

which fragment comprises two termination codons, BgIII and EcoRI sites from 5' to 3', was added.

Another clone H48-2 obtained using primers MSHNS1-2 and MSHNS1-3 was sequenced showing that said clone has no additional DNA fragment at the 5' site of No. 326 T, while has the same additional DNA fragment as that of clone H48.

[3] Modification of a DNA Fragment for the Expression of HCV Polypeptide Comprising 92 Amino Acid Sequence from No. 233 to 324 of SEQ ID NO 31, 32 in E.coli

A DNA fragment encoding a polypeptid comprising 92 amino acid sequence from Nos. 233 to 324 amino acids of SEQ ID NO 31, 32 appeared to encode an open reading frame (ORF) from HCV gene. The modification of DNA fragment was conducted in the same manner as that used for the modification of DNA fragment encoding a polypeptide of 106 amino acid sequence from amino acid Nos. 109 to 214 of SEQ ID NO 31, 32 in the above [2] except that the following primers were employed.

5' primer:

MSNS1-4: 5' GCAAGCTTATGATCGGGGGGGTCGGCAACAATAC 3' (SEQ ID NO 160) 5' primer for inserting DNA fragment into a vector containing initiation codon. MSNS1-5: 5' ATCGGGGGGGTCGGCAACAATAC 3' (SEQ ID NO 161) 3' primer:

MSNS1-6: 5' GCGAATTCAGATCTTCATCAAAGCTCTGATCTATCCCTGTCCT 3' (SEQ ID NO 162)

Each synthetic DNA was adjusted to 20 pmole/µI.

The resultant clones are H49 (primers MSNS1-4 and MSNS1-6) and H49-2 (primers MSNS1-5 and MSNS1-6).

[4] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clones N27MX24A-1, N27MX24B-1, H48-2 and H49-2 in Insect Cells

Clones N27MX24A-1 and N27MX24B-1 appears to contain an ORF which starts from the nucleotide No.2 (C). Clones H48-2 and H49-2 contain an ORF which starts from the nucleotide No.1. For the expression of polypeptide encoded by these ORF, an initiation codon ATG is inserted in frame and at an appropriate site upperstream from said gene so that the expression of the DNA might be properly effected in insect cells. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of amino acid sequence encoded by clones N27MX24A-1, N27MX24B-1, H48-2 and H49-2. When an expression vector containing an initiation codon for insect cells is used, a DNA fragment from the clone is ligated to the vector such that the frame of said DNA is in confirmity with that of the initiation codon on said vector. It also can be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of amino acid sequence encoded by clones N27MX24A-1, N27MX24B-1, H48-2 and H49-2. The modification of vector DNA was carried out by PCR. Although the modification procedures are described using clone N27MX24A-1, it can be conducted as well using clone N27MX24B-1. When insect cells were transfected with the DNA and cultivated, clones N27MX24A-1, N27MX24B-1 were expressed as in the fused form as a precursor, which was then processed, glycosylated incompletely to give a mature glycoprotein of about 70 kD accumulated intracellurarly. The modification of DNA of clones N27MX24A-1, N27MX24B-1, H48-2 and H49-2 was carried out by PCR using the following synthetic DNA as primers.

40 5' primers:

15

MS2724-4: 5' GCGTCGACGCTAGCATGCGGATCCCACAAGCCGTGGTGGAT 3' (SEQ ID NO 163) MSNS1-7: 5' GCGTCGACGCTAGCATGTTCAACGCGTCCGGATGTCCGGA 3' (SEQ ID NO 164) MSNS1-8: 5' GCGTCGACGCTAGCATGATCGGGGGGGGTCGGCAACAATAC 3' (SEQ ID NO 165) 3' primer:

45 MS2724-5: 5' GCGAATTCGCTAGCTCACTCTAAGGTGGCGTCGGCGTGGG 3' (SEQ ID NO 166) MSNS1-9: 5' GCGAATTCGCTAGCTCAACAACCGAACCAGTTGCCCTGCG 3' (SEQ ID NO 167) MSNS1-10: 5' GCGAATTCGCTAGCTCAAAGCTCTGATCTATCCCTGTCCT 3' (SEQ ID NO 168)

These three synthetic DNAs were separately adjusted to 20 pmol/ml before use.

The PCR was carried out using the same reaction solution and worked up in the same manner as described in the above [1] except that plasmid pUCHN27MX24A-1 (primers MS2724-4 and MS2724-5), pUCH48 (primers MSNS1-7 and MSNS1-9) or pUCH49 (primers MSNS1-8 and MSNS1-10) was used as a template plasmid. PCR was accomplished by repeating 10 times of reaction cycles consisting of: 1 min at 95 °C; 1 min at 50 °C and 5 min at 72 °C; and then 20 tim s of reaction cycles consisting of: 1 min at 95 °C; 1 min at 65 °C; 5 min at 72 °C. When plasmid pUCHN27MX24A-1, as a template DNA, and primers MS2724-4 and MS2724-5 were used, a desired 1268 bp DNA fragment was obtained. The other combination of plasmid pUCH48 and primers MSNS1-7 and MSNS1-9 gave a desired 352 bp DNA fragment and that of pUCH49 and primers MSNS1-8 and MSNS1-10 gave a desired 322 bp DNA fragment.

Each DNA fragment was digested with Nhel, fractionated on acrylamide gel electrophoresis and

extracted by convenional means (Molecular Cloning, Cold Spring Harbor (1982)) to obtain a DNA fragment of desired length. The resultant DNA fragment was then ligated into Nhel restriction site of a transfer v ctor pBlueBac (Invitrogen), cloned and screened for a clone which contains a single DNA fragment inserted at Nhel site. Thus, plasmids pBlueN27MX24A-1 derived from 1268 bp DNA obtained by primers MS2724-4 and MS2724-5, pBlueH48 derived from 352 bp DNA fragment obtained by primers MSNS1-7 and MSNS1-9, and pBlueH49 derived from 322 bp DNA fragment obtained by primers MSNS1-8 and MSNS1-10 were prepared.

According to the teaching shown in the protocol given by Invitrogen, the expression unit of these plasmid contains DNA fragment derived from HCV gene oriented forward and ligated to the Nhel cloning site downstream from a poyhedrin promoter.

Example 13

30

Expression of HCV Polypeptides Encoded by Clones HN27MX24A-1, HN27MX24B-1, H2N27MX24B-1, H48, H48-2, H49, and H49-2

[1] Expression of Polypeptide Encoded by Clone HN27MX24A-1, HN27MX24B-1, H48, or H49 in E.coli

Each clone encodes a part of polypeptide encoded by cDNA originated from serum of HC patient. The polypeptide encoded by each clone was expressed directly in E.coli, as it is, by subcloning said clone into an expression vector pCZ44 (Japanese Patent Publication No. 124387/1989).

A clone was digested thoroughly with restriction enzymes HindIII and BgIII, extracted with phenol/chloroform, precipitated with ethanol, separated on acrylamide gel electrophoresis. From the gel was extracted a larger DNA fragment having cohesive HindIII- and BgIII-restricted ends (Molecular Cloning, Cold Spring Harbor, 1982). The expression vector pCZ44 was digested with HindIII and BgIII. The larger fragment containing a region functional for the replication in E.coli was separated, treated in the same manner, ligated to the HindIII-BgIII fragment obtained from a clone such that the vector contains only one insertion, and cloned conventionally. The resultant plasmids were designated as plasmid pCZ2724A-1, pCZ2724B-1, pCZ48 andpCZ49 after clones HN27MX24A-1, HN27MX24B-1, H48, or H49, respectively.

Alternatively, an expression vector was constructed using an expression vector pGEX-2T (Pharmacia), in stead of pCZ44, for the expression of a fused protein between a desired polypeptide and GST. The construction was carried out substantial in accordance with the protocol taught by the manufacture (Pharmacia). Thus, the expression vector pGEX-2T was digested with BamHI. The linearized vector was ligated with a HindIII linker, and ligated with a HindIII-EcoRI DNA fragment prepared from a clone to yield expression vectors pGEX2724A-1, pGEX2724B-1, pGEX48 and pGEX49.

E.coli JM109 strain transformed with plasmid pCZ2724A-1, pCZ2724B-1, pCZ48 or pCZ49 was grown in L-Broth at 37 °C overnight (Molecular Cloning, Cold Spring Harbor, 1982). The cultured broth was diluted 50-folds by inoculating it into a freshly prepared L-Broth and the cultivation continued with shaking at 30 °C for 2 hr. At this time, IPTG was added to the culture to a final concentration of 2 mM and cultured for more than 3 hr in order to induce the expression of DNA encoding HCV-originated polypeptide by single-clone-derived transformants (E.coli JM 109 cells transformed solely by one plasmid derived from corresponding clone). Base sequences of cDNA contained in clones HN27MX24A-1 and HN27MX24B-1, and amino acid sequences deduced therefrom are shown in SEQ ID NO 31. Base sequences of cDNA contained in clones H48 and H49, and deduced amino acid sequence are shown by amino acid sequences from amino acid No. 109 to 214 and from amino acid No. 233 to 324 of in SEQ ID NO 31, respectively.

In the same manner as the above, clone HN27MX24B-1 can be used in stead of clone HN27MX24A-1 to give a polypeptide encoded by said clone. The deduced amino acid sequence of the polypeptide is shown in SEQ ID NO 32.

As mentioned in the above, plasmids pGEX2724A-1, pGEX2724B-1, pGEX48 and pGEX49 can be used to obtain transformants capable of expressing a fused protein include desired polypeptide and GST. The plasmid encodes a fused protein GST-CORE comprising GST, which has a thrombin-cleaving site at its C-terminus, and a polypeptide derived from a clone HN27MX24A-1, HN27MX24B-1, H48 or H49. Fused prot in comprising at C-terminal region a HCV polypeptide was produced in E.coli transformant transformed with either of plasmids pGEX2724A-1, pGEX2724B-1, pGEX48 and pGEX49, by culturing the cells in the same manner as that used to produce polypeptide in transformants harboring plasmid pCZ2724A-1, pCZ2724B-1, pCZ48 or pCZ49 in the presence of IPTG.

[2] Expression of Polypeptides Encoded by Clones H2N27MX24A-1, H2N27MX24B-1, H48-2, and H49-2

Polypeptides were expressed in E.coli using cDNA contained in clones H2N27MX24A-1, H2N27MX24B-1, H48-2, and H49-2 obtain d from serum of HC patient in the same manner as the above [1]. The cDNA used was that contained in clone H2N27MX24A-1, H2N27MX24B-1, H48-2, or H49-2, which clone had been previously isolated and sequenced.

Expression plasmid for each clone was constructed using pOFA (Japanese Patent Publication (KOKAI) No.84195/1990). DNA fragment from each clone was blunt-ended with T4 DNA polymerase. The expression vector pOFA was digested with KpnI and blunt-ended with T4DNA polymerase. Thus obtained DNA fragments were ligated, cloned and screened for a clone having a insertion of one DNA fragment. Thus, the desired plasmids pOFA2724A-1, pOFA2724B-1, pOFA48 and pOFA49 were prepared by subcloning a clone so that the + strand capable of expressing HCV protein should be inserted appropriately for the correct translation of said strand. It was confirmed that the cDNA from HCV was properly reconstructed by the determination of base sequence and restriction enzyme mapping of each plasmid.

Cultivation was carried out in the presence of IPTG in order to induce the expression of DNA encoding HCV-originated polypeptide by single-clone-derived transformants (E.coli JM 109 cells transformed solely by one plasmid). Base sequences of cDNAs derived from serum of a HC patient contained in clones H2N27MX24A-1, H2N27MX24B-1, H48-2 and H49-2 and amino acid sequences deduced therefrom are shown by the amino acid sequences of SEQ ID NO 31, 32, amino acid sequence from No. 109 to 214 of SEQ ID NO 31, and that from No. 233 to 324 of SEQ ID NO 31, respectively.

[3] Expression of Polypeptide Encoded by Clones N27MX24A-1, N27MX24B-1, H48-2 and H49-2 in Insect Cells

The expression of HCV-originated glycoprotein encoded by plasmid pBlueN27MX24A-1, pBlueH48 and pBlueH49 prepared in Example 12 [4] was conducted substantial in accordance with a known expression manual for baculovirus (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4, hereinafter, referred to as Maxbac, Invitrogen).

Plasmids pBlueN27MX24A-1, pBlueH48 and pBlueH49 prepared in Example 12 [4] by inserting DNA fragment containing HCV gene at the Nhel site of a transfer vector pBlueBac (Maxbac, pp.37), were recovered from E.coli host cells transformed thereby, and purified according to the method of Maniatis et al.(Molecular Cloning, Cold Spring Harbor Laboratory, pp.86 - 96 (1982)). Thus, a large amount of HCV gene-containing transfer plasmid DNA was obtained. Sf9 cells were cotransfected with 2 μg of a plasmid containing a DNA fragment from HCV gene and 1 μg of AcNPV viral DNA (Maxbac, pp.27). Sf9 cells were grown in TMN-FH medium (Invitrogen) containing 10% FCS (fetal calf serum) in a 6 cm dish until a cell density reached to about 2 x 10⁶/plate. The TMN-F medium was removed and a 0.75 ml Grace medium (Gibco) containing 10% FCS was added thereto. To the DNA mixture described in the above was added 0.75 ml of transfection buffer (attached to the kit) was thoroughly mixed by vortex and gradually added dropwise onto the Grace medium. After the culture being allowed to stand for 4 hr at 27 °C, Grace medium was replaced with 3 ml of TMN-FH medium containing 10% FCS and the dish incubated at 27 °Cfor 6 days. Three days from the incubation, there observed a few multinucleate cells and on sixth day, almost all the cells were multinuclear. The supernatant was taken into a centrifuging tube and centrifuged at 1,000 rpm, 10 min to obtain the supernatant as a cotransfected viral solution.

The cotransfected viral solution contains about 10⁸ viruses/ml and 0.5% of which were recombinant viruses. The isolation of recombinant virus was carried out by a plaque isolation method described below.

Thus, cells were adsorbed onto a 6 cm dish by seeding 1.5×10^6 cells on medium and removing the medium completely. To the dish was added $100~\mu l$ of a diluted viral solution (10^{-4} and 10^{-5} folds), separately and incubated at room temperature for 1 hr while slanting the dish every 15 min to spread the virus extensively. X-gal medium containing agarose was prepared by adding 5-bromo-4-chloro-3-indolyl- β -D-galactoside to a final concentration of $150~\mu g/l$ (Maxbac, pp.16-17) to a warm medium which had been prepared by autoclaving 2.5% baculovirus agarose (Invitrogen) at $105~^{\circ}$ C for 10 min, mixing with TMN-FH medium containing 10% FCS preheated at 46 $^{\circ}$ C at the mixing ratio of 1 : 3, and keeping the temperature at 46 $^{\circ}$ C.

After the completion of infection, virus solution was aspirated thoroughly from the dish and 4 ml of the warm X-gal medium containing agarose (previously prepared) was gently added to every dish not to peel off cells. The dish kept open by slightly sliding a lid until the agarose solidified and dried, and thereafter the dish covered, turned upside down, and incubated at 27 °C for 6 days. The plaques were observed under a phase difference microscope to find blue plaques which do not form multinucleate cells. Agarose containing blue recombinant plaques were removed with a Pasteur pipet and suspended into 1 ml of TMN-FH medium by pipetting many times. The above process which comprises: infection, 6-day incubation, and isolation of

virus containing transfer plasmid DNA is called the "plaque method". The plaque method was repeated using 100 μ I of viral suspension. After repeating said proc ss three times, there obtained a recombinant virus having a gene encoding HCV glycoprotein free from contamination with that of wild-type strain.

A viral solution of the primary recombinant virus was prepared by aspirating plaques with a Pasteur pipet, and mixing thoroughly with 1 ml of TMN-FH medium. Because the primary viral solution was low in virus density for infection, it required further treatments for concentration. Thus, $100~\mu$ l of viral solution was adsorbed onto Sf9 cells grown in a 6 cm dish to a semi-confluent, and 4 ml of TMN-FH medium was added thereto and incubated three days. The culture supernatant was recovered to yield a recombinant viral solution for infection.

For the production of HCV structural protein , a suspension of Sf9 cells in TMN-FH medium containing 10% FCS (5 x 10^6 cells/10 ml medium) was added into a 9 cm dish and kept 1 hr for adsorption. After the removal of medium, 250 μ l of recombinant viral solution was added to the dish and spread extensively. To the dish was added 10 ml TMN-FH medium containing 10% FCS and incubated at 27 °C for 4 days. The cells expressing recombinant glycoprotein of HCV were harvested by scraping up and suspended into 1,000 ml of phosphate buffered saline.

Thus, HCV-derived glycoprotein was expressed in Sf9 cells transfected with said virus.

Example 14

30

50

Identification of Expression Products as HCAg

Each expression product obtained in Example 13 was identified as HCAg because it reacted immunologically with antiserum obtained from HC patients.

Identification was conducted by Western blot technique.

E. coli cells transformed with either of expression plasmids pCZ2724A-1, pCZ2724B-1, pCZ48, pCZ49, pOFA2724A-1, pOFA2724B-1, pOFA48, pOFA49, pGEX2724A-1, pGEX2724B-1, pGEX48, pGEX49, pBlueN27MX24A-1, pBlueH48 and pBlueH49 for polypeptides encodid by clones HN27MX24A-1, HN27MX24B-1, H2N27MX24B-1, H48, H48-2, H49, and H49-2 were grown in the presence of IPTG for 3 hr or a overnight in the same manner as described in Example 13.

Recombinant strains were harvested by centrifuging 1,000 μ l of the cultured broth at 6,500 rpm, 10 min. The pellet was dissolved into a sample solution (50 mM Tris-HCl, pH6.8 containing 2% SDS, 5% mercaptoethanol, 10% glycerin, and 0.005% bromophenol blue) for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml. Sf9 cells infected with viruses which had been treated more than 3 times by plaque method were collected by scraping up and suspended into 1,000 ml of PBS and 100 μ l of the suspension was centrifuged at 6,500 rpm, 10 min to pellet the cells. The pellet was dissolved into a sample solution for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml.

The sample solutions were then boiled at 100 °C for 10 min. Ten μ I of the boiled solution was loaded onto 0.1% SDS-15% polyacrylamide gel (70 x 85 x 1 mm) together with a marker protein LMW Kit E (low-molecular weight marker protein, Pharmacia). Electrophoresis was carried out at a constant current of 30 mA for 45 min in Tris buffer (25 mM Tris, pH 8.3, 192 mM glycine, 0.1% SDS) as electrode buffer. Thereafter, DNA was transferred electophoretically to a nitrocellulose filter by superposing the gel onto a filter BA-83 (S & S), impressing a constant current of 120 mA for about 20 min between gel (cathode) and the filter (anode) as conventionally.

The transcribed filter was cut to remove a part containing a marker protein (referred to as marker filter) and that containing the sample (referred to as sample filter). The former was stained with 0.1% (w/v) amideblack 10B and the latter immersed into 0.01 M PBS (pH 7.4) containing 5% (w/v) bovine serum albumin (BSA). Serum from a HC patient was diluted 50 times with 0.01 M PBS (pH 7.4) containing 5% (w/v) BSA. To the sample filter was added 10 μ I of diluted serum and the filter allowed to stand for 2 hr at room temperature. Thereafter, the filter was washed with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3).

The sample filter was then reacted with 10 ml of horseradish peroxidase conjugated anti human IgG (Gappel) at 37 °C for 1 hr and washed with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3). The filter was th n immersed into peroxidase-color-producing solution (60 mg 4-chloro-1-naphthol, 20 ml methanol, 80 ml PBS, and 20 µl aqueous hydrogen peroxide). The colored filter was washed with distilled water and compared with the marker filter, demonstrating that colored prot in expressed by transformants transformed with plasmid pCZ2724A-1, pCZ2724B-1, pCZ48 and pCZ49 had a reasonable molecular weight as an expression product of inserted HCV gene and was identified as HCAg.

The expression product from host cells transformed with pOFA-d rived plasmids such as pOFA2724A-

1, pOFA2724B-1, pOFA48 or pOFA49 is a fused protein consisting of HCV originated polypeptide and OmpF signal peptide of E.coli and the product from host cells transformed with pGEX-d rived plasmid such as pGEX2724A-1, pGEX2724B-1, pGEX48 or pGEX49 is also a fused protein consisting of HCV originated polypeptide and GST and thrombin cleaving site wherein the latter two attached at th N-terminus of th former. Each fused protein has a reasonable molecular weight and was also identified as HCAg.

Insect cells transfected with pBlueN27MX24A-1 and pBlueN27MX24B-1, as shown in Example 13 [3], expressed HCV polypeptide encoded by clones N27MX24A-1, N27MX24B-1, H48-2, and H49-2. The expression product (M-gp70) was a glycoprotein of molecular weight of about 70 kD, which has a base sequence corresponding to the base sequence from about No.46 to about No. 395 of SEQ ID NO 31 or 32. Also glycoprotein of HCV was expressed in insect cells tranformed with plasmid pBlueH48 or pBlueH49. Thus produced glycoproteins were encoded by clones H48-2 and H49-2 and had amino acid sequences which correspond to a polypeptide having 106 amino acids from No. 109 to 214 and that of 96 amino acids from No. 233 to 324 of SEQ ID NO 31 and 32, respectively. As a result, theses glycoproteins were identified as HCAg.

Example 15

15

30

Synthesis of DNA

[1] Preparation of RNA Sample Solution

RNA sample solution was prepared by resolving the dried nucleic acid obtained in Example 1 in 30 µl of water containing 10 µl of ribonuclease inhibitor (100 U/µl, Takara Shuzo, Japan).

Oligonucleotide primers of the following base sequences were synthesized using a method well known to one of skill. Among them, antisence primers such as MS49, MS88, MS100, MS132, MS152, and MS158 were used for cloning of cDNA.

[2] Synthesis of cDNA Using Antisence Primer

To 2 μ l of RNA sample solution was added 1 μ l of 15 pmol/ μ l anti-sense primer (e.g., synthetic DNA primer such as MS158, MS152, MS132, MS49, MS88, or MS100) 2 μ l of 10 x RT buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl), 4 μ l of 25 mM MgCl₂, 8 μ l of 2.5 mM 4dNTPs, 1 μ l of water and the mixture incubated at 65 °C for 5 min then at room temperature for 5 min. To the mixture was added 1 μ l of reverse transcriptase (25 U, Life Science), 1 μ l of ribonuclease inhibitor (100 U/ μ l, Takara Shuzo) and the mixture incubated at 37 °C for 20 min, at 42 °C for 30 min, and finally at 95 °C for 2 min, which was followed by an immediate cooling to 0 °C to yield cDNA.

Amplification of DNA encoding HCAg was conducted by polymerase chain reaction (PCR) (Saiki et al., Nature 324: 126 (1986)). For the PCR, primers synthesized in the above were used as a pair of: MS48 - MS49; MS86 - MS100; MS97 - MS88; MS135 - MS132; MS155 - MS152; or MS151 - MS158.

A 100 μl mixture containing ten μl of cDNA solution, 10 μl of 10 x PCR buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 μl of 2.5 mM 4 dNTPs, 2 μl of 15 pmol/μl synthetic primer (the same primer as used in the preparation of cDNA), 3 μl of 15 pmol/μl synthetic primer (a counterpart of pairs of primers) and water was incubated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, it was mixed with 0.5 μl of Taq DNA polymerase (7 U/μl, AmpliTaq T Takara Shuzo) and overlaid with mineral oil. The resultant sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 - 55 °C for 1 min; and at 72 °C for 1 - 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). Finally, the reaction mixture was incubated at 72 °C for 7 min, which was followed by phenol/chloroform extraction and ethanol precipitation. The ethanol precipitation was carried out by adding 2.5 volumes of ethanol and either of about 1/10 volume of 3 M acetic acid or an equal volume of 4 M ammonium acetate to the aqueous solution, mixing, centrifuging at 15,000 rpm for 15 min using a rotor of about 5 cm in diameter under cooling at 4 °C to pellet the precipitates, and drying the pellet. Various amplified DNA fragments were obtained using different primers, for the cloning of cDNA and amplification thereof.

5 Example 16

Cloning and Sequencing of Amplified DNA Fragments

The cloning was carried out substantial in accordance with the method of Molecular Cloning, Cold Spring Harbor (1982).

Dried DNA fragment (at least 1 pmole) obtained in the above Example 15 [2] was blunt-ended with T4 DNA polym rase (Toyobo) and 5'-end phosphorylated with polynucleotide kinas (Toyobo) and ligated into smal site of multi-cloning sites of 5 ng to 10 ng of pUC19 cloning vector. The cloning vector had been previously treated as follows: digestion with a restriction enzyme Smal (Toyobo), phenol/chloroform-extraction, ethanol-precipitation, 5'-end dephosphorylation with alkaline phosphatase (Behringer-Manheim) (Molecular Cloning (1982) Cold Spring Harbor), phenol/chloroform extraction, and ethanol-precipitation. The ligated DNA was used to transfect into a competent E.coli JM 109 or DH5 cells (Toyobo). The transfection was carried out according to the protocol of the manufacture's instruction (COMPETENT HIGH, Toyobo). Plasmid clones were recovered from transformed cells conventionally. At least 20 transformants were obtained using pUC19 cloning vectors containing either of DNA fragments obtained using either of pairs of primers in the same manner as that described in the above Example 15 [2].

The determination of base sequence of DNA fragment was conducted by Fluorescent DNA Sequencer (GENESIS 2000, Dupont) using, as sequence primer, the following synthetic primers:

5' d(GTAAAACGACGGCCAGT)3' (SEQ ID NO 143) and

5'd(CAGGAAACAGCTATGAC)3' (SEQ ID NO 144) for the + and - strands of DNA fragment to be sequenced. Base sequences of each clone are shown in SEQ ID NO 37 - 39, 44 - 55, 103 and 104. Clones belong to the same region of HCV gene were summarized and of which sequences are shown in SEQ ID NO 33 to 36. For example, clones MX25-1, MX25-2 and MX25-3 are summarized and shown by SEQ ID NO 33.

In the same manner, clones shown by SEQ ID NO 47 to 55 are summarized as clones shown by SEQ ID NO 34 to 36.

In the Sequence Listings, base sequences shown in SEQ ID NO 33 to 39, 44 and 55 are those of + strand of HCV gene inserted into cloning vector to transfect into host cells. All the clones are double-stranded and a plasmid containing one clone and used for the sequencing of said clone is designated by adding a prefix "pUC" to the name of clone. Thus, plasmid containing clone MX25-1 is pUCMX25-1, containing MX25-2 is pUCMX25-2, containing MX25-3 is pUCMX25-3, and so on.

The base sequence shown in the Sequence Listing represents a specific sequence of cDNA corresponding to RNA isolated from serum of patient(s) suffering from HC and differs from that of cDNA obtained from RNA in serum of a healthy subject in the same manner. It was confirmed that cDNA prepared from RNA obtained from a healthy subject under more strict conditions, for instance, by repeating a reaction cycles in Example 15 [2] and [3] about 60 - 100 times (= about 3- or 4-folds), did not show any homology in base sequence with those shown in SEQ ID NO 33 to 43. Consequently, base sequences of clones shown in SEQ ID NO 33 to 43 are specific for those obtained from serum of HC patient.

The base sequences of DNA fragments were compared with a known base sequence of HCV gene. The fact that three clones MX25-1, MX25-2 and MX25-3 were obtained from serum of one HC patient in Example 9 [2] using primers MS155 and MS152 strongly suggests that there must be more than one virus in a patient.

Example 17

40

45

Preparation of Fused Clones MX25O26A-1, MX25O26B-1, N16N15A-1 and N16N15B-1, U16N15A-1 U16N15B-1, N23N15A-1,N23N15B-1, MX25N15A-1, and MX25N15B-1

[1] Preparation of Clones MX25026A-1 and MX25026B-1

One μI (about 0.5 to 1 μg/μI) each of DNA fragments from clones MX25-1 and O26-1 (prepared in Example 16) was added into a reaction mixture containing 10 μI of 10 x PCR buffer (100 mM Tris-HCI (pH8.3), 500 mM KCI, 15 mM MgCI₂, 1% gelatin), 8 μI of 2.5 mM 4 dNTPs, 5 μI each of 20 pmol/μI synthetic primers MS155 and MS158, and 76.5 μI of water. After an intimate mixing, the mixture was heated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 μI of Taq DNA polymerase (7 U/μI, AmpliTaq Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 37 °C for 1 min; and at 72 °C for 2 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by an incubation at 97 °C for 2 min. The mixture was immediately cooled to 0 °C, kept at the same temperature for 2 min, mixed with 0.5 μI of Taq DNA polymerase (7 U/μI, AmpliTaq Takara Shuzo), and overlaid with mineral oil. The sample was then treated

in the same manner as the above by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °Cfor 1 min; at 50 - 55 °C for 1 min; and at 72 °C for 2 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The amplified DNA samples were fractionated on agarose gel electrophor sis and a gel containing a desired fragment having an expected length was removed (Molecular Cloning (1982) Cold Spring Harbor) to isolate the DNA fragment therefrom conventionally. The resultant DNA fragment was then modified at the N-terminal region as described in Example 16 and ligated into Smal site of multi-cloning sites of pUC19, cloned and screened as described in Example 3 to obtain plasmids pUCMX25026A-1 and pUCMX25026B-1. CDNAs derived from serum of HC patient contained in said plasmids were designated as clones MX25026A-1 and MX25026B-1, respectively and of which base sequences are summarized in SEQ ID NO 40. Base and deduced amino acid sequences of each clone MX25026A-1 and MX25026B-1 are shown in SEQ ID NO 56 and 57. Overlapping region in clone MX25026A-1 is derived from clone MX25-1 and that of MX25026B-1 from O26-1.

In the same manner as the above, clones N16N15A-1 and N16N15B-1, and U16N15A-1 and U16N15B-1 were prepared using clones N15-1 (SEQ ID NO 39) and either of N16 (SEQ ID NO 36) and U16-4 (SEQ ID NO 37). Base sequences of clones are summarized in SEQ ID NO 41. Base and amino acid sequences of clones N16N15A-1 and N16N-15B-1 are shown in SEQ ID NO 26 and 27, respectively.

[2] Preparation of Clone N16N15-1

20

Two overlapping clones N16-1 and N15-1 were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme BstE11, clone N16-1 is cleaved at the 3' site of a nucleotide No. 576 and clone N15-1 at the 3' site of a nucleotide No.114. The ligation of A clones N16-1 and N15-1 were conducted on the basis of assumption that plasmids pUCN16-1 and pUCN15-1 contains each clone in the same orientation. As a result, a clone N16N15-1 in which clones N16-1 and N15-1 are ligated without overlapping was conducted. Thus, plasmid pUCN16-1 was digested with HindIII and BstEII to obtain a 609 bp DNA fragment comprising a HindIII-Smal DNA fragment of plasmid pUC19 attached to the 5' end of clone N16-1 (a cDNA clone from serum of a HC patient). Plasmid pUCN15-1 was digested with HindIII and BstE11 to obtain a 147 bp DNA fragment containing clone N15-1. These 609 bp and 147 bp HindIII-BstE11 fragments are then exchanged each other, cloned and screened to obtain plasmid pUCN16N15-1 containing the desired clone N16N15-1. Clones obtainable in the same manner are summarized in SEQ ID NO 41. The base and amino acid sequences of clone N16N15-1 are shown in SEQ ID NO 60.

[3] Preparation of Clones N23N15A-1 and N23N15B-1

One μI (about 0.5 to 1 μg/μI) of DNA fragment from each clone N23-1 (Example 16), and N16N15A-1, N16N15-B-1 and N16N15-1 (Example 17 [1],[2]) was added to a reaction solution containing 10 μ l of 10 xPCR buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl, 15 mM MgCl₂, 1% gelatin), 8 µl of 2.5 mM 4 dNTPs, 5 μl each of 20 pmol/μl synthetic primers MS135 and MS88, and 76.5 μl of water. After an intimate mixing, the mixture was heated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq™ Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following steps: at 95 °C for 1 min; at 37 °C for 1 min; at 72 °C for 3.5 -4 min in DNA Thermal Cycler (Parkin Elmer Cetus). After the final incubation at 92 °Cfor 2 min, the mixture was immediately cooled to 0 °C, kept at the same temperature for 2 min, mixed with 0.5 μl of Tag DNA polymerase (7 U/µI, AmpliTaq™ Takara Shuzo), and overlaid with mineral oil. The sample was then treated in the same manner as the above by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 50 - 55 °C for 1 min; and at 72 °C for 3.5 - 4 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The amplified DNA samples were fractionated on agarose gel electrophoresis and a gel containing a desired fragment having an expected length was removed (Molecular Cloning (1982) Cold Spring Harbor) to isolate the DNA fragment therefrom conventionally. The resultant DNA fragment was then modified at the N-terminal region and ligated into Smal site of the multi-cloning sites on pUC19, cloned and screened as described in Example 16 to obtain plasmids pUCN23N15A-1 and pUCN23N15B-1. cDNAs obtained from these plasmids are designated as clones N23N15A-1 and N23N15B-1, whose base and deduced amino acid sequences are shown in SEQ ID NO 61 and 62, respectively and are summarized in SEQ ID NO 42. The overlapping region in clones N23N15A-1, a fused clone of N23-1, N16N15A-1, N16N15B-1 and

N16N15A-1, is originated from clone N23-1, and that of clone N23N15B-1 is originated from clones N16N15A-1, N16N15B-1 and N16N15-1.

[4] Preparation of Clone MX25N15-1

5

Two overlapping clones MX25O26A-1 and N23N15A-1 were ligated by taking advantage of a unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme Apal, clone MX25O26A-1 was cleaved at the 3' site of C at nucleotide No. 1277 and clone N23N15A-1 at the 3' site of nucleotide No.17. A plasmid pUCMX25N15-1 in which clones MX25O26A-1 and N23N15A-1 are ligated without overlapping was constructed in the following manner by taking advantage of the fact that plasmids pUCMX25O26A and pUCN23N15A-1 contain clones MX25O26A-1 and N23N15A-1 in the same orientation at Smal site of multi-cloning sites of pUC19. Plasmid pUCMX25O26A-1 was digested with HindIII and Apal to obtain a 1310 bp DNA fragment comprising a HindIII-Smal DNA fragment of plasmid pUC19 attached to the 5' end of clone MX25O26A-1 (a cDNA clone from serum of a HC patient).

Plasmid pUCN23N15A-1 was digested with HindIII and Apal to obtain a 50 bp DNA fragment containing clone N15-1. These 1310 bp and 50 bp HindIII-Apal fragments are then exchanged each other, cloned and screened to obtain plasmid pUCMX25N15-1 containing desired clone MX25N15-1. The base and amino acid sequences of clone MX25N15-1 shown in SEQ ID NO 63.

The clones MX25O26A-1 and N23N15A-1 are ligated by PCR. The resultant base sequences are summarized in SEQ ID NO 43.

Example 18

Modification of DNA for the Expression of HCV Polypeptide Encoded by MX25N15-1

25

50

20

[1] Modification of DNA for the Expression of HCV Polypeptide Encoded by clone MX25N15-1 in E.coli

Clone MX25N15-1 appeared to contain multiple open reading frames each originated from HCV gene such as NS2 ORF (hereinafter, referred to as MK/NS2) from No. 7 (T) to 825 (G), and NS3 ORF (MK/NS3) from No. 826(G) to 2652(G) of base sequence of SEQ ID NO 43. Genes contained therein can be expressed by inserting an ATG initiation codon in frame and upperatream from said gene so that the expression thereof might be properly effected in host cells. When a partial DNA fragment derived from MK/NS2 or MK/NS3 is to be expressed, an ATG initiation codon and a termination codon are inserted upperatream and downstream from the DNA to be expressed, respectively, such that the frame of each inserted codon is in confirmity with that of the DNA. The insertion of an ATG initiation codon at the upperstream from 5' terminus of a gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of an amino acid sequence of SEQ ID NO 43. When an expression vector containing an initiation codon for E. coli. is used, a DNA fragment from the clone is ligated to the vector such that the frame of said DNA is in confirmity with that of the codon. The modification of DNA can be carried out by PCR.

For the expression of MK/NS2, the following synthetic oligonucleotide primers were used.

MSNS2-1: 5' GCAAGCTTATGTGGTTGTGGATGATGCTGCTG 3' (SEQ ID NO 169)

5' primer for the insertion of said DNA fragment into a vector having an initiation codon from a procaryotic expression vector:

MSNS2-2: 5' TGGTTGTGGATGATGCTGCTG 3' (SEQ ID NO 170) 3' primer:

MSNS2-3: 5' GCGAATTCAGATCTTCATCACCTCCGGGCGGAGACNGGNAGNCC 3' (SEQ ID NO 171) The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as described in the above according to Saiki's method in a total volume of 100 μl containing 100 ng of plasmid pUCMX25N15-1 (or plasmid pUCMX25026A-1), as a template, and 2 μl each of 3' and 5' primers MSNS2-1 and MSNS2-3. The reaction mixture was heated at 95 °C for 5 min and quench d at 0 °C. One minute later, to the mixtur was added 0.5 μl of Taq DNA polymerase (7 U/ml, AmpliTaq TM Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 25 cycles of treatments which comprises: at 95 °C for 1 minute; at 60 °C for 1 min; and at 72 °C for 3 min in DNA Thermal Cycler (Parkin Elmer Cetus). The resultant reaction solution was extracted with phenol/chloroform, and precipitated with ethanol conventionally. The amplified DNA samples were digested with HindIII and EcoRI, and fractionated on acrylamide gel electrophoresis and

extracted (Molecular Cloning, Cold Spring Harbor (1982)).

The DNA fragment is then ligated into HindIII (in case of MSNS2-2 primer, Smal site) and EcoRI sites of cloning vector pUC19 to obtain plasmid pUCHNS2-1. The base sequence of said plasmid is determined to show that it comprises a DNA fragment shown by a bas sequence from No. 7 to 825 in SEQ ID No 43 having additional DNA fragments attached to the both 5'- and 3'-termini. That is, at its 5'-terminus, the following DNA fragment comprises a HindIII restriction site followed by an initiation codon ATG was attached.

5' GCAAGCTTATG 3'

3' CGTTCGAATAC 5' (SEQ ID NO 155)

- 15 And at its 3'-terminus, the following DNA fragment comprises two termination codons, Bglll and EcoRI sites from 5' to 3' was attached.
 - 5' TGATGAAGATCTGAATTCGC 3'

3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

For an expression vector containing E.coli-derived initiation codon, DNA was modified in the same manner as the above except that primers MSNS2-2 and MSNS2-3 are employed and the amplified DNA is first blunt-ended with T4 DNA polymerase and then digested with EcoRI instead of the digestion with HindIII and EcoRI to obtain plasmid pUCH2NS2-1. The sequencing of resultant clone H2NS2-1 showed that said clone has no additional DNA fragment at the 5' terminus but, at the 3' terminus, has the same DNA fragment as that of the above clone HNS2-1.

Modification of DNA for the expression of MK/NS3 was conducted substantially in the same manner as the above except that primers MSNS3-1, 3-2 and 3-3 are used in stead of MSNS2-1, 2-2, and 2-3, respectively to obtain plasmids pUCHNS3-1 and pUCH2NS3-1, corresponding to the above plasmids pUCHNS2-1 and pUCH2NS2-1.

MSNS3-1: 5' GCAAGCTTATGGGCAACGAGNTNCTNCTIGG 3' (SEQ ID NO 172)

MSNS3-2: 5' GGCAACGAGNTNCTNCTNGG 3' (SEQ ID NO 173)

MSNS3-3: 5' GCGAATTCAGATCTTCATCACTTCAGCCGTATGAGACACTT 3' (SEQ ID NO 174)

[2] Modification of DNA for the Expression of DNA encoding HCV Polypeptide MK1 in E. coli

DNA encoding MK1 polypeptide shown by 305 amino acid sequence from No. 422 to 726 in SEQ ID NO 43 was modified in the same manner as the above [1] by inserting an initiation codon ATG in frame and the upperstream from 5' terminus of said DNA in ORF encoding MK1. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene. When an expression vector containing an initiation codon for E. coli. is used, a DNA fragment from the clone is ligated to the vector such that the frame of said DNA is in confirmity with that of the codon. The modification of DNA can be carried out by PCR using the following synthetic oligonucleotide primers.

5' primer:

MSMK1-1: 5' GCAAGCTTATGCTGTCGCCCGGGCCCATCTC 3' (SEQ ID NO 175)

50 5' primer for the insertion of a DNA fragment into a vector having an initiation codon from procaryotic expression vector:

MSMK1-2: 5' CTGTCGCCCGGGCCCATCTC 3' (SEQ ID NO 176)

3' primer:

10

20

30

MSMK1-3: 5' GCGAATTCAGATCTTCATCAACATGTGTTGCAGTCGATCAC 3' (SEQ ID NO 177)

The synthetic DNA was adjusted to 20 pmol/ml befor use.

PCR, cloning and subcloning were carried out in the same manner as described in the above [1].

Plasmid pUCMK1 was prepared by cloning a DNA fragment obtained by PCR using MSMK1-1 and MSMK1-3. The base sequence of clone MK-1 contained in plasmid pUCMK1 is determined to show that it

comprises a DNA fragment having a base sequence from No. 1264 (G) to 2178 (G) in SEQ ID No 43 having additional DNA fragments attached to the both 5'- and 3'-termini. That is, at its 5'-terminal G, a DNA fragment comprises a HindIII restriction site followed by an initiation codon ATG as follows:

5' GCAAGCTTATG 3'

3' CGTTCGAATAC 5' (SEQ ID NO 155).

And at its 3'-terminal G, a DNA fragment comprises two termination codons, <u>Bglll</u> and <u>EcoRl</u> sites from 5' to 3' as follows:

- 5' TGATGAAGATCTGAATTCGC 3'
- 3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

Another plasmid pUCMK1-2 was constructed in the same manner as the above except that primers MSMK1-2 and MSMK1-3 were employed. The sequencing of resultant clone MK1-2 showed that said clone has no additional DNA fragment at the 5' terminus but has the same additional DNA fragment as that of the above clone MK1 at its 3' terminus.

25 [3] Modification of DNA for the Expression of DNA Encoding HCV Polypeptide MK2

MK2 polypeptide shown by 322 amino acid sequence from No. 712 to 1033 in SEQ ID NO 43 appears to be HCV-derived antigenic protein which is highly reactive with antiserum from a HC patient. For the expression of DNA encoding MK2 in E.coli, said DNA was modified in the same manner as the above [2] using the following synthetic oligonucleotide primers.

5' primer:

5

10

15

20

MSMK2-1: 5' GCAAGCTTATGGGCTATACCGGNGACTTNGAC 3' (SEQ ID NO 178)

5' primer for the insertion of a DNA fragment into a vector having an initiation codon from procaryotic xpression vector:

MSMK2-2: 5' GGCTATACCGGNGACTTNGAC 3' (SEQ ID NO 179)

3' primer:

45

MSMK2-3: 5' GCGAATTCAGATCTTCAGTGCTTCGCCCAGAAGGT 3' (SEQ ID NO 180)

The synthetic DNA was adjusted to 20 pmol/ml before use. The resultant clones were designated as clone MK2 (prepared using primers MSMK2-1 and MSMK2-3) and MK2-2 (prepared using primers MSMK2-2 and MSMK2-3).

[4] Modification of DNA for the Expression of HCV Polypeptide Encoded by Clone MX25N15-1 in Insect Cells

Clone MX25N15 contains an open reading frame which starts from the nucleotide No.1 (T). For the construction of expressing plasmids for insect cells, DNA was modified essentially in the same manner as described in the above by inserting an initiation codon ATG to the upperstream from 5' terminus of said DNA in ORF. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N-terminus (amino-terminus) of the total or a part of amino acid sequence which is encoded by clone MK25N15. When an expression vector containing an initiation codon for insect cells is used, a DNA fragment from the clone is ligated to the vector such that the expression of said DNA can be initiated at the codon. In this case, the modification may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N-terminus (amino-terminus) of the total or a part of amino acid sequence which is encoded by clone MK25N15. When an insect cell transformed with an expression plasmid containing MK25N15, said clone was expressed as a precursor polypeptide having an amino acid sequence, which at least contains amino acids from No. 167 to 502, which was then processed, glycosylated and accumulated intracellurarly.

The modification of clone MX25N15 DNA was carried out by PCR employing the following synthetic oligonucleotides as primers.

5' primer:

MS2515-1: 5' GCGCTAGCATGTGGTTGTGGATGATGCTG 3' (SEQ ID NO 181)

3' primer:

MS2515-2: 5' GCGAATTCGCTAGCTCACAGCCGGTTCATCCACTGCAC 3' (SEQ ID NO 182)

The synthetic DNA was adjusted to 20 pmol/ml before use.

The PCR was carried out using the same reaction solution and worked up in the same manner as described in the above except that plasmid pUCMX25N15-1 was used as a template plasmid, primers MS2515-1 and MS2515-2 were used as primers, and the PCR was carried out repeating 10 times the following reaction cycles consisting of: 1 min at 95 °C; 1 min at 50 °C and 5 min at 72 °C; and then repeating 20 times of the following reaction cycle consisting of: 1 min at 95 °C; 1 min at 65 °C; 5 min at 72 °C to obtain a desired 3586 bp DNA fragment.

The amplified DNA samples were digested with Nhel and fractionated on acrylamide gel electrophoresis and extracted as conventionally (Molecular Cloning, Cold Spring Harbor (1982)). The DNA fragment was then ligated into Nhel restriction site of a transfer vector pBlueBac (Invitrogen), cloned and screened for a clone containing a single DNA insert at Nhel site conventionally to obtain a plasmid pBlueMX25N15-1.

Taking account of the instruction provided by the manufacture (Invitrogen), the expression unit of the resultant plasmid contains a DNA fragment derived from HCV gene oriented forward and ligated to the Nhel cloning site down stream from a poyhedrin promoter.

Example 19

25

Expression of HCV Polypeptides Encoded by MX25N15, and Polypeptides MK/NS2, MK/NS3, MK1 and MK3

[1] Expression of Polypeptide Encoded by Clones HNS2-1, HNS3-1, MK1 or MK2

Clones HNS2-1, HNS3-1, MK1 and MK2 encode polypeptide fragments derived from a polypeptide encoded by cDNA obtained from a serum of a HC patient and were expressed as it is in E.coli. Construction of expression vector for each clone was carried out by subcloning it into an expression vector pCZ44 (Japanese Patent Publication No. 1-124387/1989).

Each clone was digested thoroughly with restriction enzymes HindIII and BgIII, extracted with phenol/chloroform, precipitated with ethanol, separated on acrylamide gel electrophoresis, and extracted a DNA fragment having cohesive HindIII- and BgIII-restricted ends from the gel (Molecular Cloning, Cold Spring Harbor, 1982). The expression vector pCZ44 was digested with HindIII and BgIII and the larger fragment containing functional region for expressing DNA was separated and treated in the same manner. The both DNA fragments were ligated at their HindIII and BgIII sites and cloned. The resultant plasmids were named plasmids pCZHNS2-1, pCZHNS3-1, pCZMK1 and pCZMK2 after clones HNS2-1, HNS3-1, MK1 and MK2, respectively.

Alternatively, expression vectors encoding polypeptides encoded by clones HNS2-1, HNS3-1, MK1 and MK2 were constructed using an expression vector pGEX-2T (Pharmacia) designed to express a fused protein of desired peptide and β -glutathione-S-transferase (GST). The construction was carried out substantial in accordance with the protocol taught by the manufacture (Pharmacia).

The expression vector pGEX-2T was digested with BamHI. To the linearized vector was ligated a HindIII linker to obtain a DNA fragment having EcoRI and HindIII restriction sites at its 3'- and 5'-termini. Each clone was digested with HindIII and EcoRI to obtain DNA fragments encoding desired HCV polypeptides. The two fragments were then ligated at their HindIII and EcoRI sites to obtain expression vectors pGEXHNS2-1, pGEXHNS3-1, pGEXMK1, and pGEXMK2, respectively. These plasmids contain DNAs encoding GST, thrombin-cleaving sequence, and desired clone from upperstream to downstream.

E.coli JM109 strain was transformed with a plasmid pCZHNS2-1, pCZHNS3-1, pCZMK1 or pCZMK2 and transformant was grown in L-Broth at 37 °C conventionally (Molecular Cloning, Cold Spring Harbor, 1982). The cultured broth was inoculated into a fresh L-Broth to decrease the concentration to 1/50 and cultured with shaking at 30 °C for 2 hr. IPTG (isopropyl-β-D-galactopyranoside) was added to the culture to a final concentration of 2 mM in order to induce exclusiv ly the expression of DNA by single-clone-derived transformant (E.coli transformed with a single plasmid) and cultivation continued for more than 3 hr. Thus, the transformant produced a polypeptide encoded by the clone. Deduced amino acid sequences of polypeptides encoded by cDNA derived from clones HNS2-1, HNS3-1, MK1 and MK2 are shown by amino

acid sequences from 422 to 726 and 712 to 1033 in SEQ ID NO 43.

E.coli JM 109 cells were transformed with expression v ctor pGEXHNS2-1, pGEXHNS3-1, pGEXMK1, or pGEXMK2 and cultured in the same manner as the above. The expression of gene encoding a fused polypeptide was induced by IPTG. The resultant fused protein comprises GST, a thrombin-cleaving site at its C-terminus, and a polypeptide derived from a clone HNS2-1, HNS3-1, MK1 or MK2.

[2] Expression of Polypeptides Encoded by Clones H2NS2-1, H2NS3-1, MK1-2, and MK2-2

Clones H2NS2-1, H2NS3-1, MK1-2, and MK2-2, which had been isolated and sequenced, were expressed in E.coli to give an fused protein using a substantially the same manner as the above.

The fused protein comprises, for instance, a signal peptide of OmpF, an outer membrane protein of E.coli, and a polypeptide encoded by either of the above-mentioned clones can be expressed using, as the expression vector, pOFA (Japanese Patent Publication No. 84195/1990). DNA fragment from each clone H2NS2-1, H2NS3-1, MK1-2, or MK2-2 was blunt-ended with T4DNA polymerase. The expression vector pOFA was digested with Kpnl and blunt-ended with T4DNA polymerase. Thus obtained DNA fragments were ligated, cloned and screened for a clone having a insertion of one DNA fragment. Thus, the desired plasmids pOFANS2-1, pOFANS3-1, pOFAMK1, and pOFAMK2 were prepared by subcloning a clone so that the + strand responsible for the expression of HCV protein should be inserted appropriately for the correct translation of the clone. It was confirmed by the determination of base sequence and mapping of each plasmid that the HCV-derived cDNA was reconstructed properly.

E.coli JM109 cells were transformed with an expression vector obtained in the above and induced the expression of DNA by growing host cells under the presence of IPTG as previously described. Transformants expressed a fused protein of a signal peptide of OmpF and a HCV polypeptide encoded by each clone. DNA sequences and deduced amino acid sequences of polypeptides encoded by clones HNS2-1, HNS3-1, MK1 and MK2 are shown by amino acid sequences from 422 to 726 and 712 to 1033 in SEQ ID NO 43. Thus, according to this method, HCV polypeptide was expressed as a fused protein between OmpF signal peptide and polypeptide encoded by each clone.

[3] Expression of Polypeptide Encoded by Clone MX25N15 in Insect Cells

30

35

The expression of HCV polypeptide encoded by plasmid pBlueMX25N25 prepared in Example 18, [4] was conducted substantial in accordance with a known expression manual for baculovirus (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4, hereinafter, referred to as Maxbac, Invitrogen).

Plasmid pBlueMX25N15, a plasmid prepared by inserting DNA fragment containing HCV gene at the Nhel site of a transfer vector pBlueBac (Maxbac, pp.37) was recovered from E.coli/pBlueMX25N15, and purified according to the method of Maniatis et al.(Molecular Cloning, Cold Spring Harbor Laboratory, pp.86 - 96 (1982)) to obtain a large amount of HCV gene-containing transfer plasmid DNA. Sf9 cells were cotransfected with 2 μg of plasmid pBlueMX25N15 and 1 μg of AcNPV viral DNA (Maxbac, pp.27). Thus, Sf9 cells were grown in TMN-FH medium (Invitrogen) containing 10% FCS (fetal calf serum) in 6 cm dish until a cell density reached to about 2 x 10⁶/plate. The TMN-F medium was removed and a 0.75 ml Grace medium (Gibco) containing 10% FCS was added thereto. A DNA solution of 2 μg plasmid pBlueMX25N15 DNA and 1 μg AcNPV viral DNA in 0.75 ml of transfection buffer (attached to the kit) was thoroughly mixed by vortex and gradually added dropwise onto the Grace medium. After allowing to stand for 4 hr at 27 °C, the Grace medium was replaced with 3 ml of TMN-FH medium containing 10% FCS and the dish incubated at 27 °C for 6 days. Three days from the incubation, there observed a few multinucleate cells and on sixth day, almost all the cells were multinuclear. The supernatant was taken into a centrifuging tube and centrifuged at 1,000 rpm, 10 min to obtain the supernatant as a cotransfected viral solution.

The cotransfected viral solution contains about 10⁸ virus/ml and 0.5% of which were recombinant virus. The isolation of recombinant virus was carried out by plaque isolation method described below.

Thus, cells were adsorbed onto a 6 cm dish by seeding 1.5×10^6 cells on medium and removing the medium. To the dish was added $100 \,\mu$ l of a diluted viral solution (10^{-4} and 10^{-5} folds), separately and incubated at room temperature for 1 hr while slanting the dish every 15 min to spread th virus extensively. X-gal medium containing agarose was prepared by adding 5-bromo-4-chloro-3-indolyl- β -D-galactoside to a final concentration of 150 μ g/l to a warm medium which had been prepared by autoclaving 2.5% baculovirus agarose (Invitrogen) at 105 °C for 10 min, mixing with TMN-FH medium containing 10% FCS preheated at 46 °C at the mixing ration of 1 : 3, and keeping the temperature at 46 °C.

After the completion of infection, virus solution was aspirated thoroughly from the dish and 4 ml of the

warm X-gal medium containing agarose (previously prepared) was gently added to every dish not to peel off cells. The dish kept open by slightly sliding a lid until the agarose solidified and dried, and thereafter th dish covered, turned upside down, and incubated at 27 °C for 6 days. The plaques were observed under a phase difference microscope to find blue plaques which do not form multinucleate c lls. Agarose containing blue recombinant plaques were removed by an aspirating pipet and suspended into 1 ml of TMN-FH medium by pipetting many times. The above process which comprises: infection, 6-day incubation, and isolation of virus containing transfer plasmid DNA is called the "plaque method". The plaque method was repeated using 100 μ l of viral suspension. After repeating said process three times, there obtained a recombinant virus having a gene encoding HCV glucoprotein free from contamination with wild-type strain.

A viral solution of the primary recombinant virus was prepared by aspirating plaques with a Pasteur pipet, and mixing thoroughly with 1 ml of TMN-FH medium. Because the primary viral solution was low in virus density for infection, it was further treated. Thus, 100 µl of viral solution was adsorbed onto Sf9 cells grown in a petri dish (6 cm in diameter) to a semi-confluent, and 4 ml of TMN-FH medium was added thereto and incubated three days. The culture supernatant was recovered to yield a recombinant viral solution for infection.

2. Infection of Sf9 Cells with Recombinant Viral Solution

A suspension of Sf9 cells in TMN-FH medium containing 10% FCS (5 x 10^6 cells/10 ml medium) were added into a Petri dish (9 cm, in diameter) and kept 1 hr for adsorption. After the removal of medium, 250 μ l of recombinant viral solution was added to the dish and spread extensively. To the dish was added 10 ml TMN-FH medium containing 10% FCS and incubated at 27 μ C for 4 days. The cells expressing recombinant glycoprotein of HCV were harvested by scraping up and suspended into 1,000 ml of phosphate buffered saline. Thus, HCV glycoproteins were expressed by Sf9 cells infected with said viral solution.

Example 20

30

10

Identification of Expression Product as HCAg

Each expression product obtained in Example 19 was identified as HCAg because it reacted immunologically with antiserum obtained from HC patient. Identification of the expression product as HCAg was conducted by Western blot as follows. E. coli cells were transformed with either of plasmids described in Example 19 [1], [2], such as plasmids pCZHNS2-1, pCZHNS3-1, pCZMK1, pCZMK2 pGEXHNS2-1, pGEXHNS3-1, pGEXMK1, pGEXMK2, pOFANS2-1, pOFANS3-1, pOFAMK1, and pOFAMK2 for polypeptides encoded by clones HNS2-1, HNS3-1, MK1, MK2 HNS2-1, HNS3-1, MK1, MK2, H2NS2-1, H2NS3-1, MK1-2, and MK2-2, respectively and grown under the presence of IPTG for 3 hr or a overnight.

Recombinant strains were harvested by centrifuging 1,000 µl of the cultured broth at 6,500 rpm, 10 min. The pellet was dissolved into a sample solution (50 mM Tris-HCl, pH6.8 containing 2% SDS, 5% mercaptoethanol, 10% glycerin, and 0.005% bromophenol blue) for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml. The sample solution was then boiled at 100 °C for 10 min. Ten µl of the boiled solution was loaded on 0.1% SDS-15% polyacrylamide gel (70 x 85 x 1 mm) together with a marker protein LMW Kit E (low-molecular weight marker protein, Pharmacia). Electrophoresis was carried out at a constant current of 30 mA for 45 min in Tris buffer (25 mM Tris, pH 8.3, 192 mM glycine, 0.1% SDS) as electrode buffer. Thereafter, DNA was transferred electophoretically to a nitrocellulose filter by superposing the gel onto a filter BA-83 (S & S), impressing a constant current of 120 mA for about 20 min between gel (cathode) and the filter (anode) as conventionally.

The transcribed filter was cut to separate a region containing a marker protein (marker filter) and that containing the sample (sample filter) and the former was stained with 0.1% (w/v) amideblack 10B and the latter immersed into 0.01 M PBS (pH 7.4) containing 5% (w/v) bovine serum albumin (BSA). A serum from a patient suffering from hepatitis C was diluted 50 times with 0.01 M PBS (pH 7.4) containing 5% (w/v) BSA. To the sample filter was added 10 μ I of diluted serum and the filter allowed to stand for 2 hr at room temperature. Thereafter, the filter was wash d with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3).

The sample filter was then reacted with 10 ml of horseradish peroxidase conjugated anti human IgG (Gappel) at 37 °C for 1 hr and washed with PBS containing 0.1% (v/v) Tween 20 for 20 min (x3). The filter was then immersed into peroxidase-color-producing solution (60 mg 4-chloro-1-naphthol, 20 ml methanol, 80 ml PBS, and 20 µl aqueous hydrogen peroxide). The colored filter was washed with distilled water and compared with the marker filter to demonstrate that the product d veloped color had a reasonable

molecular weight as an expression product of HCV gene contained in expression plasmid pCZHNS2-1, pCZHNS3-1. pCZMK1 or pCZMK2, and was identified as HC-associated antigens.

The expression product from host cells transformed with plasmid pOFANS2-1, pOFANS3-1, pOFAMK1 or pOFAMK2 is a fused prot in consisting of HCV originated polyp ptide and OmpF signal peptide of E.coli wherein the latter two attached at the N-terminus of the former. The expression product from host cells transformed with plasmid pGEXHNS2-1, pGEXHNS3-1, pGEXMK1 or pGEXMK2 is also a fused protein consisting of HCV originated polypeptide and GST and thrombin cleaving site wherein the latter attached at the N-terminus of the former. These fused proteins have reasonable molecular weight and was also identified as HC-associated antigens.

Example 21

10

15

20

25

35

50

55

Synthesis of cDNA

In the present Example 21, water used is ultra-pure water which was preparaed by autoclaving (x 2) distilled water.

[1] Preparation of RNA Sample Solution

RNA isolated in Example 1 was dried, dissolved into 0.3 M (pH 7.0) sodium acetate, treated with phenol/chloroform (x 1), mixed with 2.5 volumes of ethanol, and centrifuged (15000 rpm, 20 min, at room temperature) with a rotor of about 5 cm in diameter to yield a pellet of nucleic acid. The pellet was then dried and dissolved into 30 μ l of water containing 10 μ l of ribonuclease inhibitor (100 U/ μ l, Takara Shuzo, Japan) to give a nucleic acid solution, which was then subjected to the cDNA synthesis.

[2] Synthesis of cDNA Using Antisense Primer

To 2 μ l of RNA sample solution prepared in above [1] was added 1 μ l of 15 pmol/ μ l anti-sense primer selected from a group of synthesized primers MS126, MS119, MS161, MS162, MS121, and MS163 shown in Table, 2 μ l of 10 x RT buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl), 4 μ l of 25 mM MgCl₂, 8 μ l of 2.5 mM 4dNTPs, 1 μ l of water and the mixture incubated at 65 - 70 °C for 5 min then at room temperature for 5 min. To the mixture was added 1 μ l of reverse transcriptase (25 U, Life Science), 1 μ l of ribonuclease inhibitor (100 U/ μ l, Takara Shuzo) and the mixture incubated at 37 °C for 20 min, at 42 °C for 30 min, and finally at 95 °C for 2 min, which was followed by an immediate cooling to 0 °C (synthesis of cDNA).

Amplification of DNA containing specific sequences was conducted by PCR (Saiki et al., Nature 324: 126 (1986)). Thus, 100 µl mixture containing ten µl of cDNA solution, 10 µl of 10 x PCR buffer (100 mM Tris-HCI (pH8.3), 500 mM KCl, 15 mM MgCl2, 1% gelatin), 8 µI of 2.5 mM 4 dNTPs, 2 µI of 15 pmol/µI synthetic DNA primer (the same primer as used in the synthesis of cDNA), 3 µI of 15 pmol/µI synthetic DNA primer (a counterpart of pair of primers, i.e., MS127-MS126, MS118-MS119, MS159-MS161, MS160-MS162, MS120-MS163, or MS120-MS121) and water was incubated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, it was mixed with 0.5 µl of Taq DNA polymerase (7 U/µl, AmpliTaq™ Takara Shuzo) and overlaid with mineral oil. The resultant sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 40 - 55 °C for 1 min; and at 72 °C for 1 - 5 min in DNA Thermal Cycler (Parkin Elmer Cetus). Finally, the reaction mixture was incubated at 72 °C for 7 min, which was followed by phenol/chloroform extraction and ethanol precipitation to obtain different amplified DNA fragments derived from either of above-mentioned pairs of primers. The ethanol precipitation was carried out by adding 2.5 volumes of ethanol and either of about 1/10 volume of 3 M acetic acid or an equal volume of 4 M ammonium acetate to the aqueous solution, mixing, centrifuging at 15,000 rpm for 15 min using a rotor of about 5 cm in diameter under cooling at 4 °C to pellet the precipitates, and drying the pellet.

Example 22

Cloning and Sequencing of Amplified DNA Fragments

The cloning was carried out substantial in accordance with the method of Molecular Cloning, Cold Spring Harbor (1982).

Dried DNA fragment (at least 1 pmole) obtained in the above Example 21, [2] was blunt-ended with T4

DNA polymerase (Toyobo) and 5'-end phosphorylated with polynucleotide kinase (Toyobo) and ligated into smal site of multi-cloning sites of 5 ng to 10 ng of pUC19 cloning v ctor. The cloning vector had b en previously treated as follows: digestion with a restriction enzyme Smal (Toyobo), phenol/chloroform extraction, ethanol pr cipitation, 5'-end dephosphorylation with alkalin phosphatas (Behringer-Manheim) (Molecular Cloning (1982) Cold Spring Harbor), ph nol/chloroform extraction, and ethanol-precipitation. The ligated DNA was used to transfect into a competent E.coli JM 109 or DH5 cells (Toyobo). The transfection was carried out according to the protocol of the manufacture's instruction (COMPETENT HIGH, Toyobo). Plasmid clones were recovered from transformed cells conventionally. At least 20 transformants were obtained using pUC19 cloning vectors containing either of DNA fragments obtained using either of pairs of primers in the same manner as that described in Example 21, [2].

Plasmid DNA was isolated from corresponding transformant by an usual method and sequenced. The determination of base sequence was conducted by means of Fluorescent DNA Sequencer (GENESIS 2000, Dupont) using, as sequence primer, the following synthetic primers:

5' d(GTAAAACGACGGCCAGT)3' (SEQ ID NO 143) and

5'd(CAGGAAACAGCTATGAC)3' (SEQ ID NO 144) for the + and - strands of DNA fragment to be sequenced. When the DNA fragment is longer than about 200 bp, the determination was conducted by subcloning said DNA into a clone of dilation mutant in order to make sure the sequencing.

DNA fragment obtained using either of pairs of primers shown in Example 21 and whose base sequence was determined is listed below.

Pair of primers	cione(s)
MS127-MS126	N22-1, N22-3, N22-4, H22-3, H22-8, H22-9
MS118-MS119	N17-1, N17-2, N17-3, H17-1, H17-3
MS159-MS161	O28-1, O28-2, 028-4
MS160-MS162	N29-1, N29-2, N29-3
MS120-MS163	O30-2, O30-3, O30-4
MS120-MS121	N18-2, N18-3, N18-4, H18-1, H18-2, H18-3

The alphabet letter used to express each clone represents the serum of HC patient used in Example 1. The base sequence of clones proved to have a homology with a known base sequence of HCV gene. The region on HCV gene corresponding to each clone was designated as follows.

region on HCV gene
N22
N17
O28
N29
O30
N18

Among resultant clones, base and amino acid sequences of clones N22-1, N17-3, O28-1, N29-1, N18-4, O30-3 are shown in SEQ ID NO 76, 81, 86, 89, 92, and 98, respectively. Base sequences of other clones obtained in the same manner are listed below in alignment with a base sequence of a clone which disclosed in Seq. Lis. In the list, the base sequence of a clone disclosed in the Seq. Lis. is given at the uppermost column, which is follwed by others in the same region, showing only the bases which are different from those of the clone to be referred to (that shown in the uppermost column). The figure following the name of clone represents the nucleotide number of the base at 5' terminus of the sequence. The nucleotide is numbered from 5' terminus (base No. 1) conventionally.

55

20

25

35

	BASE SEQUENCE	OF EACH	CLONE IN N22 REGION
	N22-1.NUC	1 :	GGCATGTGGGCCCAGGGGAGGGGGGCTGTGCAGTGGATGAACCGGCTGATA
5	(SEQ ID NO: 7	6)	
	N22-3.NUC	1:	••••••
	(SEQ ID NO: 7	7) .	
10	H22-3.NUC	1:	••••••
	(SEQ ID NO:	78)	•
	H22-8.NUC	1:	•••••
15	(SEQ ID NO: 7	9)	·
	H22-9.NUC	1:	•••••
	(SEQ ID NO: 8	0)	
20	N22-1.NUC	51 :	${\tt GCGTTTGCTTCGCGGGGCAACCATGTCTCCCCCACGCACTATGTGCCTGA}$
	N22-3.NUC	51 :	••••••
	H22-3.NUC	51 :	·····C············T···················
25	H22-8.NUC	51 :	•••••C••C••••••••••
	H22-9.NUC	51 :	·····C···········T·······C············
	N22-1.NUC	101 :	AAGCGACGCCGCAGCGCGCGCCACCCAGATCCTCCCAACCTTACCATCA
30 .	N22-3.NUC	101 :	•••••••••••••••••••••••••••••••••••••••
	H22-3.NUC	101 :	G
	H22-8.NUC	101 :	G·····································
35	H22-9.NUC	101 :	G
	N22-1.NUC	151 :	CTÇAGCTGTTGAAGAGGCTTCACCAGTGGATTAATGAGGACTGCTCCACG
	N22-3.NUC	151 :	·····T································
40	H22-3.NUC	151 :	••••••••••••••••••••••••••••••••••••••
	H22-8.NUC	151 :	••••••••••••••••••••••••••••••••••••••

	H22-9.NUC	151 : ··································
	N22-1.NUC	201 : CCATGCTCCGGCTCGTGGCTCAGGGATGTTTGGGACTGGATATGCACGGT
5	N22-3.NUC	201 :
	H22-3.NUC	201 : ·····T··T··T···
	H22-8.NUC	201 : ·····T··T··T····
10	H22-9.NUC	201 : ·····T··T··T····
	N22-1.NUC	251 : ATTGGCTGATTGCAAGACCTGGCTCCAGTCCAAGCTCCTGCCGCGGTTAC
	N22-3.NUC	251 : ···································
15	H22-3.NUC	251 : G···AG···C·T······
	H22-8.NUC	251 : G···AG···C·T····························
	H22-9.NUC	251 : G···AG···C·T····························
20	N22-1.NUC	301 : CGGGGGTCCCTTTTTCTCATGCCAGCGTGGGTACAAGGGGGTTTGGCGG
	N22-3.NUC	301 : ···································
	H22-3.NUC	301 : ····A······CC······A······A···C·····
25	H22-8.NUC	301 : ····A······CC·T······A··········A···C·····
	H22-9.NUC	301 : ····A······CC······A·····A····
	N22-1.NUC	351 : GGAGATGGCATCATGTATACCACCTGCCCATGTGGAGCACAAATCACCGG
30	N22-3.NUC	351 : ·····C······
	H22-3.NUC	351 : ···································
	H22-8.NUC	351 : ···································
35	H22-9.NUC	351 : •••••••••••••••••••••••••••••••••••
	N22-1.NUC	401 : ACATGTCAAAAACGGTTCTATGAGGATCGTTGGGCCTAGAACCTGTAGCA
40	N22-3.NUC	401 : ·····
40	H22-3.NUC	401 : ···································
	H22-8.NUC	401 :

5

	H22-9.NUC	401 :	······································
	N22-1.NUC	451 :	${\tt ACACGTGGCACGGAACATTTCCCATCAACGCGTACACCACAGGCCCCTGC}$
5	N22-3.NUC	451 :	· · · · · · · · · · · · · · · · · · ·
	H22-3.NUC	451 :	······GC
	H22-8.NUC	451 :	······G··C······
10	H22-9.NUC	451 :	······································
	N22-1.NUC	501 :	${\tt ACACCCTCCCGGCGCCAAACTATTCCAGGGCGTTGTGGCGGGTGGCCAT}$
	N22-3.NUC	501 :	· · · · · · · · · · · · · · · · · · ·
15	H22-3.NUC	501 :	······································
	H22-8.NUC	501 :	······································
20	H22-9.NUC	501 :	••••••••••••••••••••••••••••••••••••••
20	N22-1.NUC	551 :	${\tt TGAGGAGTATGTGGAGGTCACGCGGGTGGGGGGATTTCCACTACGTGACGG}$
	N22-3.NUC	551 :	
25	H22-3.NUC	551 :	
	H22-8.NUC	551 :	
	H22-9.NUC	551 :	
30	N22-1.NUC	601 :	${\tt GCATGACCACTGACAACGTGAAATGCCCATGCCAGGTTCCGGCCCCCGAA}$
	N22-3.NUC	601 :	
	H22-3.NUC	601 :	
35	H22-8.NUC	601 :	С
	H22-9.NUC	601 :	
	N22-1.NUC	651 :	TTCTTCACAGAATTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTG
40	N22-3.NUC	651 :	
	H22-3.NUC	651 :	AAA
	H22-8.NUC	651 :	······G···G···G······A······A·····A·····

	H22-9.NUC	651 :	••T••••G••G•••••••A••••A•••••
	N22-1.NUC	701 :	${\tt CAAACCTCTCCTGCGGGATGAGGTCACATTCCAGGTCGGGCTCAACCAAT}$
5	N22-3.NUC	701 :	•••••
	H22-3.NUC	701 :	······A······
	H22-8.NUC	701 :	······A······
10	H22-9.NUC	701 :	······A······
	N22-1.NUC	751 :	ATACGGTTGGGTCACAGCTCCCATGTGAGCCCGAACCGGATGTAACAGTG
	N22-3.NUC	751 :	·····G···
15	H22-3.NUC	751 :	TCC···································
	H22-8.NUC	751 :	TCCC
	H22-9.NUC	751 :	TCC·····G····A····C·····G····
20	N22-1.NUC	801 :	GTCACCTCCATGCTCACC
	N22-3.NUC	801 :	•••••
	H22-3.NUC	801 :	•••••
25	H22-8.NUC	801 :	
	H22-9.NUC	801 :	•••••
	BASE SEQUENCE	OF EACH	CLONE IN N17 REGION
30	N17-3.NUC	1 :	TGTGAGCCCGAACCGGATGTAACAGTGGTCACCTCCATGCTCACCGACCC
	(SEQ ID NO:	81)	
	N17-1.NUC	1 :	
35	(SEQ ID NO:	82)	
	N17-2.NUC	1:	•••••••••••••••••
	(SEQ ID NO:	83)	
40	H17-1.NUC	1 :	······································
	(SEQ ID NO:	84)	•

	H17-3.NUC	1	:	······································
	(SEQ İD NO:	85)		
5	N17-3.NUC	51	:	$\tt CTCCCACATTACAGCAGAGGGGGCTAGGCGTAGGCTGACCAGAGGGTCTC$
	N17-1.NUC	51	:	
	N17-2.NUC	51	:	
10	H17-1.NUC	51	:	· · · · · · · · · · · · · · · · · · ·
	H17-3.NUC	51	:	
	N17-3.NUC	101	:	${\tt CCCCTTCCTCGACCAGTTCTTCAGCTAGTCAGTTGTCTGCGCTTTCTTCG}$
15	N17-1.NUC	101	:	•T•••••CA••••T•
	N17-2.NUC	101	:	•T•••••CA••••T•
	H17-1.NUC	101	:	CT.GC
20	H17-3.NUC	101	:	······································
	N17-3.NUC	151	:	${\tt CAGGCAACATGCACTACCCATCAGGGCGCCCCAGACACTGACCTCATCGA}$
	N17-1.NUC	151	:	$\textbf{A} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \textbf{G} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \boldsymbol{\cdot} \cdot$
25	N17-2.NUC	151	:	AG
	H17-1.NUC	151	:	AGT.A.TGG
	H17-3.NUC	151	:	AGT.A.TGG
30	N17-3.NUC	201	:	GGCCAACCTCCTGTGGCGGCAGAGATGGGCGGAAACATCACCCGCGTGG
	N17-1.NUC	201	:	
	N17-2.NUC	201	:	
35	H17-1.NUC	201	:	AGT
	H17-3.NUC	201	:	A
	N17-3.NUC	251	:	AGTCAGAGAACAAGATAGTAATTCTAGACTCTTTTGAACCGCTTCGAGCG
40	N17-1.NUC	251	:	·····C······
	N17-2.NUC	251	:	···T······

	H17-1.NUC	251 :	· · · · · · · · · · · · · · · · · · ·
	H17-3.NUC	251 :	······································
5	N17-3.NUC	301 :	GAGGAGGATGA
	N17-1.NUC	301 :	
	N17-2.NUC	301 :	•••••
10	H17-1.NUC	301 :	•••••
	H17-3.NUC	301 :	•••••
	BASE SEQUENCE	OF EACH	CLONE IN 028 REGION
15	028-1.NUC	1:	${\tt GTGGTAGTCCTGGACTCGTTGGAGCCGCTTCAAGCGAAGGAAG$
	(SEQ ID NO:	86)	·
	028-2.NUC	1:	······G·····A·····
20	(SEQ ID NO:	87)	
	O28-4.NUC	1:	······································
	(SEQ ID NO:	88)	
25	028-1.NUC	51 :	GGAAGTGTCCGTTGCGGCGGAGATCCTGCGGAAGACCAGGAAATTCCCCG
	O28-2.NUC	51 :	·····A·····A·····A·····A·····
	028-4.NUC	51 :	
30	028-1.NUC	101. :	CAGCGATGCCCGTATGGGCACGCCCGGACTACAACCCACCATTACTAGAG
	O28-2.NUC	101 :	***************************************
	028-4.NUC	101 :	
35	028-1.NUC		TCTTGGAAGAACCCGGACTACGTCCCTCCAGTGGTACACGGGTGCCCATT
	028-2.NUC	151 :	· · · · · · · · · · · · · · · · · · ·
	028-4.NUC	151 :	G
40	028-1.NUC	201 :	GCCGCCTACCAAGGCCCCTCCAATACCACCTCCACGAAGAAAGA
	028-2.NUC	201 :	······································

	028-4.NUC	201	:	······································
	028-1.NUC	251	:	TTGTCCTGACAGAATCCTCCGTGTCCTCTGCCTTGGCGGAGCTTGCTACA
5	028-2.NUC	251	:	····C············A····················
	028-4.NUC	251	:	
	028-1.NUC	301	:	${\tt AAGACCTTTGGCAGTTCCGGATCGTCGGCCGTCGACAGCGGCACGGCGAC}$
10	028-2.NUC	301	:	
	028-4.NUC	301	:	
	028-1.NUC	351	:	$\tt CGGCCCTCCTGACCAGGCCTCCGCCGAAGGAGGATGCAGGATCCGACGCTG$
15	028-2.NUC	351	:	T
	028-4 NUC	351	:	
	028-1.NUC	401	:	${\tt AGTCGTACTCCTCCATGCCCCCCTTGAGGGAGAGCCGGGGACCCCGAT}$
20	028-2.NUC	401	:	T
	028-4.NUC	401	:	······································
	028-1.NUC	451	:	${\tt CTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCCAGCGAGGACGTCTAGGGAGGG$
25	028-2.NUC	451	:	
	028-4.NUC	451	:	
	028-1.NUC	501	:	${\tt CGTCTGCTGGATGTCCTACACATGGACAGGCGCCTTAATTACACCAT}$
30	028-2.NUC	501	:	
	028-4.NUC	501	:	
	028-1.NUC	551	:	GCGCCGCGGAGGAGCAAGCTGCCCATTAATGCGCTGAGCAACCCTTTG
35	028-2.NUC	551	:	Т
	028-4.NUC	551	:	АТ
	028-1.NUC			CTGCGCCACCACAACATGGTCTATGCCACAACATCCCGCAGCGCAAGCCA
40	028-2.NUC			•••••••••••••
	028-4.NUC	601	:	т

	028-1.NUC	651 :	${\tt GCGGCAGAAAAAGGTCACATTTGACAGACTGCAAGTCCTGGATGACCACT}$
	028-2.NUC	651 :	•••••
5	028-4.NUC	651 :	•••••
	028-1.NUC	701 :	ACCGGGACGTGCTCAAGGACATGAAGGCCAAGGCGTCCAC
	028-2.NUC	701 :	••••••
10	028-4.NUC	701 :	•••••
	BASE SEQUENCE	OF EACH	CLONE IN N29 REGION
	N29-1.NUC	1:	ACTACCGGGACGTGCTGAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAG
15	(SEQ ID NO:	89)	
	N29-2.NUC	1:	
	(SEQ ID NO:	90)	
20	N29-3.NUC	1:	••••••••••••
	(SEQ ID NO:	91)	
	N29-1.NUC	51 :	GCTAAACTTCTATCTGTAGAGGAAGCCTGCAAGCTGACGCCCCCACACTC
25	N29-2.NUC	51 :	T
	N29-3.NUC	51 :	······································
	N29-1.NUC	101 :	GGCCAGATCTAAATTTGGCTACGGGGCAAAGGACGTCCGGAGCCTGTCCA
30	N29-2.NUC	101 :	•••••••••••
	N29-3.NUC	101 :	· · · · · · · · · · · · · · · · · · ·
	N29-1.NUC	151 :	GCAAGGCCGTTAACCACATCCGCTCCGTGTGGAAGGACTTGCTGGAAGAC
35	N29-2.NUC	151 :	••••••••••
	N29-3.NUC	151 :	······································
	N29-1.NUC		ACTGAGACACCAATTGACACCACCATCATGGCAAAAAATGAGGTTTTCTG
40	N29-2.NUC	201 :	••••••••••
	N29-3.NUC	201 :	A

	N29-1.NUC	251 :	TGTTCAACCAGAGAAAGGAGGCCGCAAGCCAGCTCGCCTTATCGTATTCC
	N29-2.NUC	251 :	
5	N29-3.NUC	251 :	
	N29-1.NUC	301 :	CAGACTTGGGGGTTCGTGTGCGAGAAAATGGCCCTCTACGACGTGGTC
	N29-2.NUC	301 :	
10	N29-3.NUC	301 :	••••••••••••
	N29-1.NUC	351 :	TCCACTCTTCCTCAGGCCGTGATGGGCTCCTCATACGGATTCCAGTACTC
	N29-2.NUC	351 :	
15	N29-3.NUC	351 :	••••••
	N29-1.NUC	401 :	CCCTGGACAGCGGGTCGAGTTCCTGGTGAATGCCTGGAAGTCAAAGAAGA
	N29-2.NUC	401 :	•••••••••••
20	N29-3.NUC	401 :	
	N29-1.NUC	451 :	GCCCTATGGGCTTTGCATATGACACCCGCTGTTTTGACTCAACGGTCACC
	N29-2.NUC		т
25	N29-3.NUC	451 :	т
	N29-1.NUC	501 :	GAGAACGACATCCGT
	N29-2.NUC	501 :	• • • • • • • • • • • • • • • • • • • •
30	N29-3.NUC	501 :	•••
	BASE SEQUENCE	OF EACH	CLONE IN 030 REGION
	030-3.NUC	1 :	TGGGGATCCCGTATGATACCCGCTGCTTTGACTCAACGGTCACTGAGAAT
35	(SEQ ID NO:	98)	
	030-2.NUC	1 :	•••••••••••
	(SEQ ID NO:	99)	
40	O30-4.NUC	1 :	
	(SEQ ID NO:	100)	•

	030-3.NUC	51	:	GACATCCGTGTCGAGGAGTCAATTTACCAATGTTGTGACTTGGCCCCCGA
	030-2.NUC	50	:	
5	030-4.NUC	51	:	
	030-3.NUC	101	:	${\tt GGCCAGACAGGCCATAAGGTCACTCACAGAGCGGCTTTACATCGGGGGCCCCCCCC$
	030-2.NUC	100	:	· · · · · · · · · · · · · · · · · · ·
10	030-4.NUC	101	:	
	030-3.NUC	151	:	CCCTGACTAATTCAAAGGGGCAGAACTGCGGTTATCGCCGGTGCCGCGTC
	030-2.NUC	150	:	·····A······C·························
15	030-4.NUC	151	:	······································
	030-3.NUC	201	:	${\tt AGCGGCGTGCTGACGACTAGCTGCGGTAATACCCTCACATGTTACTTGAA}$
	030-2.NUC	200	:	
20	030-4.NUC	201	:	
	030-3.NUC	251	:	GGCCTCTGCAGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTTG
	030-2.NUC	250	:	
25	030-4.NUC	251	:	•••••••••••••••••••••••••••••••••••••••
	030-3.NUC	301	:	TGTGCGGAGACGACCTTGTCGTTATCTGTGATAGCGCGGGAACTCAGGAG
	030-2.NUC	300	:	А
30	030-4.NUC	301	:	А
	030-3.NUC	351	:	GACGCGGCGAGCCTACGAGTCTTCACGGAGGCTATGACTAGGTACTCTGC
	030-2.NUC	350	:	
35	030-4.NUC	351	:	
	030-3.NUC			CCCCCCGGGGACCCGCCCCAACCAGAATACGACTTGGAGCTGATAACAT
	030-2.NUC			
40	030-4.NUC	401	:	
	030 3 1970	4 5 1	_	

	O30-2.NUC	450 : ····C·······C·······················
	030-4.NUC	451 :C
5	O30-3.NUC	501 : TACTATCTCACCCGTGACCCCACCACCCCCCTAGCGGGGCTGCGTGGGA
	O30-2.NUC	500 :C
	030-4.NUC	501 : ···································
10	030-3.NUC	551 : GACAGCTAGACACTCCAGTCAACTCCTGGCTAGGCAACATCATCATGT
	O30-2.NUC	550 :
	030-4.NUC	551 :
15	O30-3.NUC	601 : ACGCGCCCACCTTATGGGCAAGGATGATTCTGATGACCCACTTCTTCTCC
	O30-2.NUC	600 : ·T······
	030-4.NUC	601 :
20	O30-3.NUC	651 : ATCCTTCTAGCCCAGGAGCAACTTGAAAAAGCCCTAGATTGTCAGATCT
	030-2.NUC	650 :
	030-4.NUC	651 :
25	030-3.NUC	701 : CGGGGCCACTTACTCCATTGAGCCACTTGACCTACCTCAGATCATTCAAC
	030-2.NUC	700 : T
	030-4.NUC	701 :
30	030-3.NUC	751 : GACTCCACGGTCTTAGCGCATTTTCACTCCATAGTTACTCTCCAGGTGAC
	030-2.NUC	750 : ·······
	030-4.NUC	751 :тт
35	030-3.NUC	801 : ATCANTAGGGTGGCTTCATGCCTCAGGAAACTTGGGGTACCGCCCTTGCC
	030-2.NUC	800 :
	030-4.NUC	801 :
40	030-3.NUC	851 : AGTCTGGAGACATCGGGCCAGAAGCGTCCGCGCTAAGCTACTGTCCCAG
	030-2.NUC	850 :

	030-4.NUC	851 : •	
	030-3.NUC		GGGGAGGCCGCCACCTGTGGCAAATACCTCTTCAACTGGGCAGTAAAG
5	030-2.NUC		
	030-4.NUC	901 :	
	030-3.NUC		ACCAAGCTCAAACTCACTCCAATCCCAGAAGCGTCCCAGCTGGACTTGTC
10	030-2.NUC		
	030-4.NUC		
	030-3.NUC		CGGCTGGTTCGTTGCTGGTTACAGCGGGGGAGACATATATCACAGCCTGT
15	030-2.NUC		••••••••••••
	030-4.NUC		••••••
	030-3.NUC		CTCGTGCCCGACCCCGCTGGTTCATGTGGTGCCTACTCCTACTTTCCGTA
20	030-2.NUC		·····
	030-4.NUC		
	O30-3.NUC		GGGGTAGGCATCTACCTGCTCCCCAACCGATGAGCGGGGAGCTAAACACT
25	030-2.NUC		
	030-4.NUC	1101 :	
	O30-3.NUC		CCAGGCCAATAGGCCATCCCC
30	030-2.NUC	1150 :	
	030-4.NUC	1151 :	
	BASE SEQUENCE		CLONE IN N18 REGION
35	N18-4.NUC	1 :	TGGGGATCCCGTATGATACCCGCTGCTTTGACTCAACGGTCACTGAGAAT
	(SEQ ID NO:	92)	
	N18-2.NUC	1 :	A
40	(SEQ ID NO:		·
	N18-3 NIIC	1:	

	(SEQ ID NO:	94)		
	H18-1.NUC	1	:	·····A·····G·
5	(SEQ ID NO:	95)		
	H18-2.NUC	1	:	A
	(SEQ ID NO:	96)		
10	H18-3.NUC	1	:	ACG.
	(SEQ ID NO:	•		
	N18-4.NUC			GACATCCGTACTGAGGAGTCAATTTATCAATGTTGTGACTTGGACCCCGA
15	N18-2.NUC			······································
	N18-3.NUC	51	. :	
	H18-1.NUC	51	. :	··T·····GT·······C··C··C···············
20	H18-2.NUC			··T·····GT······C···C···C··············
	H18-3.NUC	51	. :	TGT
	N18-4.NUC			GGCCAGACAGGCCATAAGGTCGCTCACAGAGCGGCTTTATATCGGGGGCC
25	N18-2.NUC			
	N18-3.NUC			••••••
	H18-1.NUC			
30	H18-2.NUC			••••••••••••••••
	H18-3.NUC	10	l :	······································
	N18-4.NUC			CCTTGACCAATTCAAAAGGGCAAAACTGCGGCTATCGCCGGTGCCGCCC
35	N18-2.NUC			•••••
	N18-3.NUC			·····T································
	H18-1.NUC			· ··C····T············G·······T······T······-T·
40	H18-2.NUC			- · · C · · · · T · · · · · · · · · · · ·
	H18-3.NUC	15	1	: ••C••••T•••••••••G••••••T•••••••••

	N18-4.NUC	201 : AGCGGCGTGCTGACGACTAGCTGCGGTAATACCCTCACATGTTACTTGAA
	N18-2.NUC	201 :
5	N18-3.NUC	201 : ······
	H18-1.NUC	201 : ···································
	H18-2.NUC	201 :T
10	H18-3.NUC	201 : ···································
	N18-4.NUC	251 : GGCCTCTGCAGCCTGTCGAGCTGCGAAGCTCCAGGACTGCACGATGCTCG
	N18-2.NUC	251 : ······G······G·······
15	N18-3.NUC	251 :
	H18-1.NUC	251 : ·······A······A······
	H18-2.NUC	251 : ······A······A······
20	H18-3.NUC	251 : ·····A·····A·····
	N18-4.NUC	301 : TGTGCGGAGACGACCTTGTCGTTATCTGTGAAAGCGCGGGAACCCAGGAG
	N18-2.NUC	301 :G
25	N18-3.NUC	301 :
	H18-1.NUC	301 :G
	H18-2.NUC	301 :G
30	H18-3.NUC	301 :G
	N18-4.NUC	351 : GACGCGGCAAACCTACGAGTCTTCACGGAGGCTATGACCAGGAATTCCGC
	N18-2.NUC	351 :G
35	N18-3.NUC	351 :
	H18-1.NUC	351 :G
	H18-2.NUC	351 : ••••••G•••••••
40	H18-3.NUC	351 : ······G·······
	N18-4.NUC	401 : C
45		N10 2 NVG 401 -
		N18-2.NUC 401 : •
		N18-3.NUC 401: •
50		H18-1.NUC 401 : •
		H18-2.NUC 401 : •
		H18-3.NUC 401 : •

Bas sequences of clones in each of six regions are summarized in SEQ ID NO 64 to 69. Base sequences of SEQ ID NO 64 to 69, 76 - 100 show the base sequences of + - strand of DNA fragments which were derived from HCV gene and inserted into each plasmid used for the transformation. These clones are double stranded DNA. Plasmids used for the sequencing of clones were designated by adding a

prefix "pUC" to the name of each clone, for example, plasmid used for sequencing the clone N22-1 was designated as plasmid pUCN22-1. Each plasmid contained one DNA molecule.

These base sequences represents those of clones obtained by cloning the cDNA synthesized from RNA isolated from serum of patient(s) suffering from HC. Therefore, these sequences are specific for clones originated from serum of HCV-infected patients but can not be found or obtained from serum of healthy subjects. Thus, cDNA prepared from RNA (if there are any) obtained from a healthy subject under more strict conditions, for instance, by increasing (3 or 4 folds) the reaction cycles of PCR in Example 21 [2], by repeating them 60 - 100 times, did not show any homology in base sequence with those shown in SEQ ID NO 64 - 69, and 76 - 100. Consequently, base sequences of clones shown in SEQ ID NO 64 - 69, and 76 - 100 are specific for those obtained from serum of HC patient.

The above table indicates that there must be more than one virus in a patient.

Example 23

Preparation of Clone 1530U

[1] Preparation of Clones 1728, 2217, and 2918

Clones N17-3 and O28-1 were ligated using overlapping region by PCR. One µI (about 0.5 to 1 µg/µI) of each DNA fragment from clones N17-3 and O28-1 (311 and 740 bp, respectively) was added into a reaction mixture containing 10 µl of 10 x PCR buffer (100 mM Tris-HCl (pH8.3), 500 mM KCl, 15 mM MgCl2, 1% gelatin), 8 µl of 2.5 mM 4 dNTPs, 5 µl each of 20 pmol/µl synthetic primers MS118 and MS161, and 76.5 μl of water. After an intimate mixing, the mixture was heated at 95 °C for 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 µI of Taq DNA polymerase (7 U/µI, AmpliTaq™ Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 37 °C for 1 min; and at 72 °C for 2 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by an incubation at 97 °C for 2 min. The mixture was immediately cooled to 0 °C, kept at the same temperature for 2 min, mixed with 0.5 μl of Taq DNA polymerase (7 U/μl, AmpliTaq™ Takara Shuzo), and overlaid with mineral oil. The sample was then treated in the same manner as the above by repeating 25 times of reaction cycle, which comprises the following treatments: at 95 °Cfor 1 min; at 55 °C for 1 min; and at 72 °C for 2 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The two DNA fragments were ligated and amplified by PCR. The ligated DNA sample was fractionated on agarose gel electrophoresis and a gel containing about 1000 bp fragment was excised from the gel (Molecular Cloning (1982) Cold Spring Harbor). The resultant DNA fragment was then modified as described in Example 22 and ligated into Small site of multi-cloning sites of pUC19, cloned and screened as described in Example 22 to obtain plasmid pUC1728. The resultant clone derived from serum of HC patient was designated as clone 1728 and whose base sequence is given in SEQ ID NO 8.

In the same manner as the above, plasmid pUC2217 was obtained from clones N22-1 and N17-3, which plasmid contains at Smal site a DNA fragment derived from serum of HC patient in the following order from 5' to 3' site: EcoRI restriction site from pUC19, DNA from clone N22-1, DNA from clone N17-3, and HindIII restriction site. Base and amino acid sequences of clone 2217 are given in SEQ ID NO 70.

In the same manner as the above, clone 2918 was obtained from clones N29-1 and N18-4 whose base and amino acid sequences are given in SEQ ID NO 72.

[2] Preparation of Clone 1718

There is a 43 bp sequence common to clones 1728 and 2918. These fragments of 1004 and 857 bp were ligated by PCR substantial in accordance with the procedures as those described in the above [1] except that the elongation step in PCR reaction using Taq polymerase was conducted at 72 °C for 5 min. The resultant plasmid pUC1718 contained a DNA fragment having a base sequence derived from HCV gene at Smal site in which EcoRI site of pUC19 is located to the 5' sit of clone N17-3. (N17 region is located to 5' site of N18 region on HCV gene). Base and amino acid sequences of clone 1718 is given in SEQ ID NO 73.

[3] Preparation of Clone 2218

Overlapping clones 2217 and 2218 were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzym Xbal, pUC2217 was cleaved at two sites, i.e., in a sequence from clone N17-3 and the other in a sequence from pUC19, and a small fragment of less than about 40 bp and a large fragment containing most of the sequences from vector and clone 2217 were separated on agarose gel electrophoresis and the larger fragment, pUC2217/Xbal, was extracted. Plasmid pUC1718 was also cleaved at two sites within a sequence from clone N17-3 and one from pUC19, and a larger DNA fragment 1718/Xbal of about 1545 bp containing most of the sequences from vector and clone 1718 was separated on agarose gel electrophoresis and extracted. The ligation of clones 2217 and 1718 was accomplished on the basis of an assumption that plasmids pUC2217 and pUC1718 contain each DNA fragment in the same orientation. Thus, 10 ng of pUC2217/Xbal and 50 ng of 1718/Xbal was ligated in the presence of T4DNA ligase and the ligation mixture was incubated with competent E.coli JM109 cells and cloned in the same manner as Example 22. Transformants containing plasmid pUC2218 which contains clone 17-3 religated at Xbal site. The plasmid pUC2218 contains at its Smal site, EcoRI site and the following regions without overlapping: clones N22-1, N17-3, O28-1, N29-1, N18-4. Base and amino acid sequences of the resultant clone 2218 is given in SEQ ID NO 74

[4] Ligation of N15 Region and O30 Region Corresponding to 3' Terminal Region of HCV Gene

Clone O30-3 is shown in SEQ ID NO 98. Plasmid pUCO30 contains a DNA fragment having a sequence corresponding to 3' terminal region of HCV gene at Smal site of pUC19 in the order of, from 5' to 3', EcoRl site and clone O30-3. Plasmid pUCN15 contains a DNA fragment of HCV gene, clone N15, forwardly at Smal site of pUC19 in the order of, from 5' to 3', EcoRl site and clone N15.

Plasmid pUCO30 was cleaved at a cloning site, Sacl, of pUC19 and blunt ended with T4 DNA polymerase conventionally, which was followed by the cleavage at another cloning site, Hindlll, of pUC19 to obtain a DNA fragment O30 (SacT4/Hind) derived from HCV gene. Plasmid pUCN15 was digested with Xbal, blunt ended, and digested with Hindlll to obtain a larger DNA fragment pUCN15 (XbaT4/Hind) which contains a sequence from clone N15-1 and all the region of Hindlll fragment of pUC19. About 80 ng of DNA fragment O30 (SacT4/Hind) and about 20 ng of DNA fragment pUCN15 (XbaT4/Hind) were ligated in the presence of T4DNA ligase in 20 μl of reaction mixture. The ligation mixture was incubated with COMPETENT HIGH JM109 (Toyobo) according to the protocol provided by the manufacture and transformants containing desired plasmid pUC15-30 were isolated. Taking advantage of the fact that said plasmid pUC15-30 has only one site which can be cleaved by restriction enzymes Bglll and Hindlll, it was subjected to PCR using a primer MS174 having a Bglll site in sequence derived from clone O30-3.

PCR was conducted using, as a template, 10 ng of pUC1530 and primers MS174 and MS175. PCR fragment was then digested with BgIII and HindIII and the resultant fragment ligated to a BgIII-HindIII fragment containing the vector fragment of pUCO30 to obtain plasmid pUC15-30U having polyU attached to the 3' terminus of clone 30-3.

[5] Ligation of N15 to O30 Regions

25

There is an Apal site within a region common to N15 and N22 regions. There also is an Apal site within a region common to N18 and O30 regions. A DNA fragment isolated from pUC2218 with Apal was inserted into Apal site of pUC15-30 appropriately to obtain plasmid pUC1530U. Thus, plasmid pUC2218 was digested with Apal and 30 ng of desired DNA fragment, pUC2218/Apal, was isolated by agarose gel electrophoresis conventionally. Plasmid pUC15-30 was digested with Apal completely and desired DNA fragment was isolated and dephosphorylated. Ligation was conducted using 30 ng of pUC2218/Apal and 20 ng of dephosphorylated DNA fragment in a final volume of 10 µl. All the ligation mixture was used to transform COMPETENT HIGH JM109 (Toyobo). From transformants, desired plasmid pUC1530U which contains at the cloning site, Smal, a clone 1530U having a sequence from regions N15 to O30 without overlapping was prepared. Base and amino acid sequences of clone 1530 were determined in the same manner as that used in Example 22 and shown in SEQ ID NO 75.

The amino acid sequ nce of the ligated region comprising N15 to O30 regions has a high homology with a part of non-structural protein NS4 and NS5 of Flavivirus, a related strain of HCV. It was also confirmed that said region is homologous to a sequence encoding a part of NS4 region and all of the NS5 region by comparison with a known sequence of entire HCV gene disclosed by aforementioned Chiron, Shimotohno, or Takamizawa. As a conclusion, clones herein disclosed and whose sequence are shown in Seq Lis correspond to a part of NS4 and all of NS5. As the next step, polypeptides encoded by said clone

was evaluated as to the ability to react immunologically with antiserum of HC patients.

Example 24

Modification of DNA for the Expression of HCV Polypeptide Encoded by Clone 1530U

The expression of all or a part of regions of clone 1530U which encodes HCV polypeptide can be accomplished using any of methods which will be hereinafter described.

[1] Modification of DNA for the Expression of a part of HCV Polypeptide Encoded by Clone 1530U in E.coli

This method is used to express a desired polypeptide free from additional amino acid sequence.

Clone 1530U appears to encode an ORF derived from HCV gene (hereinafter, referred to as NS5N) from No.1246 (C) to 1692 (C) of base sequence of SEQ ID NO 75, which can be expressed by inserting an ATG initiation codon at 5' site of said gene in frame. When a part of amino acid sequence of NS5N is desired to be expressed, ATG initiation codon and termination codon were inserted to 5' and 3' site of a gene encoding said amino acid sequence such that the frame of these codons are in confirmity with that of the gene. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of an amino acid sequence of SEQ ID NO 75. This may happen when a sequence of pUC19 is inserted between ATG codon and DNA encoding HCV polypeptide at the time of insertion of ATG codon. The modification of DNA was carried out by PCR using the following synthetic DNAs as primer. 5' primer:

MSNS5-1: 5' GCAAGCTTATGCAGCGTGGGTACAAGGGGGTT 3' (SEQ ID NO 183)

3' primer:

45

50

55

MSNS5-2: 5' GCGAATTCAGATCTTCATCAGAGCTGTGACCCAACCGTATATTGGTT 3' (SEQ ID NO 184)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out according to Saiki's method in a total volume of 100 µl containing 100 ng of plasmid pUC2217 (or pUCN22-1 which contains the same region), and 2 µl each of the above 3' and 5' primers. The reaction mixture was heated at 95 °C for 5 min and quenched at 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/ml, AmpliTaq ™ Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 25 cycles of treatments which comprises: at 95 °C for 1 minute; at 55 °C for 1 min; and at 72 °C for 1 min in DNA Thermal Cycler (Parkin Elmer Cetus). The resultant reaction solution was extracted with phenol/chloroform and precipitated with ethanol conventionally. The amplified DNA samples were digested with HindIII and EcoRI and fractionated on acrylamide gel electrophoresis and the desired DNA fragment was extracted. The resultant DNA fragment was then ligated into HindIII and EcoRI sites of a cloning vector pUC19, cloned and screened to obtain plasmid pUCNS5N, which was then sequenced. The clone NS5N has a modified base sequence of that from No.1246 (C) to 1692 (C) of SEQ ID NO 75, wherein, at the 5' site of said sequence, the following DNA fragment:

- 5' GCAAGCTTATG 3'
- 3' CGTTCGAATAC 5' (SEQ ID NO 155)

which comprises a HindIII restriction site followed by an initiation codon ATG was added, and, at the 3' site of said sequence, the following DNA fragment:

- 5' TGATGAAGATCTGAATTCGC 3'
- 3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

which comprises two termination codons, BgIII and EcoRI sites from 5' to 3' was added.

[2] Modification of DNA for the Expression of HCV Polypeptide Encoded by MKCNS5 Region in Insect Cells

MKCNS5 region is an ORF derived from HCV gene encoding an amino acid sequence from No. 415 to No. 1411 of SEQ ID NO 75. For the expression of polypeptide, an initiation codon ATG is insert d at 5' site of said gene in frame so that the expression of the gene might be properly eff cted in insect cells. The insertion of an ATG initiation codon at the upperstream from 5' terminus of said gene may be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of all or a part of the amino acid sequence encoded by HCV gene. When an expression vector containing an initiation codon for insect cells is used, a DNA fragment from the clone is ligated to the vector such that the frame of said DNA is in confirmity with that of the initiation codon on said vector. It also can be accompanied by an addition of a foreign polypeptide which is not encoded by HCV gene to the N' terminus (amino terminus) of all or a part of the amino acid sequence encoded by HCV gene. Polypeptides encoded by MKCNS5 was expressed in insect cells as a single precursor polypeptide subject to that said polypeptide comprises, at least, the amino acid sequence from No. 415 to 1411 of SEQ ID NO 75, which precursor was then processed by, for example, glycosylation and accumulated intracellurarly. The modification of DNA of clone MKCNS5 region was carried out by PCR using the following synthetic DNA as primers. 5' primers:

MKCNS5-1: 5' GCGCTAGCATGGGGTACAAGGGGGTTTGGCGGG 3' (SEQ ID NO 185) 3' primer:

MKCNS5-2: 5' GCGCTAGCTCATCGGTTGGGGAGCAGGTAGAT 3' (SEQ ID NO 186)

These primers were designed to introduce Nhel site at both ends of said gene in order to insert said gene into Nhel site of transfer vector pBlueBac (Invitrogen). Therefore, the use of these primers are not critical and others can be used which are designed for introducing said gene into any other transfer vectors for insect cells. The above two synthetic DNAs were adjusted to 20 pmol/ml before use.

The PCR was carried out using the same reaction solution and worked up in the same manner as described in the above [1] except that primers MKCNS5-1 and MKCNS5-2 and, as a template plasmid, 20 ng of plasmid pUC1530U were used. PCR was accomplished by repeating 10 times of reaction cycles consisting of: 1 min at 95 °C; 1 min at 50 °C and 5 min at 72 °C; and then 20 times of reaction cycles consisting of: 1 min at 95 °C; 1 min at 65 °C; 5 min at 72 °C to yield a desired 3013 bp DNA fragment.

The DNA fragment was digested with Nhel, fractionated on acrylamide gel electrophoresis and a DNA fragment of desired length was extracted. The resultant DNA fragment was then ligated into Nhel site of a transfer vector pBlueBac (Invitrogen), cloned and screened for a clone which contains a single DNA fragment inserted at Nhel site to obtain plasmid pBlueMKCNS5.

According to the teaching shown in the protocol given by Invitrogen, the expression unit of said plasmid contains DNA fragment derived from HCV gene oriented forward and ligated to the Nhel cloning site downstream from a poyhedrin promoter.

Example 25

25

30

45

Expression of HCV Polypeptides Encoded by Clones NS5N, MKCNS4bNS5 in E.coli

Each clone encodes a part of polypeptide encoded by cDNA originated from serum of HC patient. The polypeptide encoded by each clone was expressed in E.coli, as it is, by subcloning said clone into an expression vector pCZ44 (Japanese Patent Publication (KOKAI) No. 124387/1989).

A DNA fragment having a sequence of clone NS5N obtained in Example 24 was digested thoroughly with restriction enzymes HindIII and BgIII, extracted with phenol/chloroform, precipitated with ethanol, separated on acrylamide gel electrophoresis. From the gel was extracted a DNA fragment having cohesive HindIII- and BgIII-restricted ends. The expression vector pCZ44 was digested with HindIII and BgIII. The larger fragment containing a region functional for the expression of DNA was separated, treated in the same manner, ligated to the HindIII-BgIII fragment obtained from a clone and cloned conventionally. The resultant plasmid was designated as plasmid pCZNS5N after the clone.

Alternatively, expression vectors were constructed using an expression vector pGEX-2T (Pharmacia) designed to xpress a fused protein substantial in accordance with the protocol taught by the manufacture (Pharmacia). The expression vector pGEX-2T was digested with BamHI. To the linearized vector was ligated a HindIII linker to obtain a DNA fragment having EcoRI and HindIII restriction sites at its 3'- and 5'-termini. Each clone was digested with HindIII and EcoRI to obtain DNA fragments encoding desired HCV polypeptides. The two fragments were then ligated at their HindIII and EcoRI sites such that the frame of the codon is in confirmity with the amino acid of the clone.

For example, the following region corresponding to HCV polypeptide (hereinafter, referred to as clone MKCNS4bNS5) having a 863 amino acid s quence from No. 306 to 1168 of SEQ ID NO 75 was expr ssed in E.coli. A DNA fragment encodes MKCNS4bNS5 is named as clone MKCNS4bNS5.

The above region appears to be a HCAg which can immunologically react with antiserum from HC patients in high efficiency. This region can be expressed using pCZ44 for the construction of expression vector. However, it also can be expressed as a fused polypeptide with GST.

Plasmid pUC2218 (2 ng) was digested thoroughly with restriction enzymes HindIII, PvuII and SspI and separated on acrylamide gel electrophoresis. From the gel was extracted about 200 ng of DNA fragment containing a region from clone 2218, which fragment was then blunt-ended. The DNA fragment 2218 (Hin/Pvu/T4) was inserted into HindIII site of pGEXH10 which has a modified sequence of pGEX-2T, wherein the sequence between BamHI and EcoRI sites of pGEX-2T is changed as follows:

5' GGATCCCCCCAAGCTTGGGGGAATTC 3'

BamHI HindIII EcoRI (SEQ ID NO 187)

The expression vector pGEXH10 (1 ng) was digested with HindIII completely and blunt-ended. DNA fragment from pGEXH10 (20 ng) was ligated to 50 ng of DNA fragment 2218 (Hin/Pvu/T4), transformed, and cloned conventionally. The resultant plasmid pGEX2218 encodes a fused polypeptide comprising GST linked to the N22 region of DNA fragment 2218 (Hin/Pvu/T4).

E.coli JM109 strain transformed with plasmid pCZNS5N was grown in L-Broth at 37 °C overnight (Molecular Cloning, Cold Spring Harbor, 1982). The cultured broth was diluted 50-folds by inoculating it into a freshly prepared L-Broth and the cultivation continued with shaking at 30 °Cfor 2 hr. At this time, IPTG was added to the culture to a final concentration of 2 mM in order to induce the expression of DNA encoding HCV-originated polypeptide by single-clone-derived transformants (E.coli JM 109 cells transformed solely by said plasmid). Deduced amino acid sequence of cDNA derived from clone NS5N corresponds to that of No. 1246 to 1692 of SEQ ID NO 75.

In the same manner as the above, plasmid pGEX2218 can be used to express a fused protein between polypeptide MKCNS4bNS5 and GST. The plasmid , as instructed by Pharmacia, contains a sequence encoding a region specifically cleaved by thrombin at C-terminal region of GST, followed by a sequence of clone 2218 (it also contains a short sequence derived from pUC19). The fused protein can be expressed in the same manner used for the expression of HCV polypeptide encoded by plasmid pCZNS5. Thus, E.coli transformants transformed with plasmid pGEX2218 were grown in the presence of IPTG.

Example 26

40

15

Expression of MKCNS5 Region in Insect Cells

The expression of HCV-originated protein encoded by plasmid pBlueMKCNS5 prepared in Example 24 [2] was conducted substantial in accordance with a known expression manual for baculovirus (MAXBAC™ BACULOVIRUS EXPRESSION SYSTEM MANUAL VERSION 1.4, hereinafter, referred to as Maxbac, Invitrogen).

Plasmid pBlueMKCNS5 prepared in Example 24 [2] by inserting DNA fragment containing HCV gene at the Nhel site of a transfer vector pBlueBac (Maxbac, pp.37), was recovered from E.coli host cells transformed thereby, and purified according to the method of Maniatis et al (Molecular Cloning, Cold Spring Harbor Laboratory, pp.86 - 96 (1982)). Thus, a large amount of HCV gene-containing transfer plasmid DNA was obtained. Sf9 cells were cotransfected with 2 µg of a plasmid containing a DNA fragment from HCV gene and 1 µg of AcNPV viral DNA (Maxbac, pp.27). Sf9 cells were grown in TMN-FH medium (Invitrogen) containing 10% FCS (fetal calf serum) in a Petri dish (6 cm diameter) until a cell density reached to about 2 x 10⁶/plate. The TMN-F medium was removed and a 0.75 ml Grace medium (Gibco) containing 10% FCS was added th reto. To th DNA mixture described in the abov was added 0.75 ml of transfection buffer (attached to the kit) was thoroughly mixed by vortex and gradually added dropwise onto the Grace medium. After the culture being allowed to stand for 4 hr at 27 °C, Grace medium was replaced with 3 ml of TMN-FH medium containing 10% FCS and the dish incubated at 27 °Cfor 6 days. Three days from the incubation, there observed a few multinucleate cells and on sixth day, almost all the cells were multinuclear. The supernatant was taken into a centrifuging tube and centrifuged at 1,000 rpm, 10 min to obtain the

supernatant as a cotransfected viral solution.

The cotransfected viral solution contains about 108 viruses/ml and 0.5% of which were r combinant viruses. The isolation of recombinant virus was carried out by a plaque isolation method described below.

Thus, cells were adsorbed onto a Petri dishes (6 cm diameter) by seeding 1.5×10^6 cells on medium and removing the medium completely. To the dish was added $100 \, \mu l$ of a diluted viral solution (10^{-4} and 10^{-5} folds), separately and incubated at room temperature for 1 hr while slanting the dish every 15 min to spread the virus extensively. X-gal medium containing agarose was prepared by adding 5-bromo-4-chloro-3-indolyl- β -D-galactoside to a final concentration of $150 \, \mu g/l$ to a warm medium which had been prepared by autoclaving 2.5% baculovirus agarose (Invitrogen) at $105 \, ^{\circ}\text{C}$ for 10 min, mixing with TMN-FH medium containing 10% FCS preheated at $46 \, ^{\circ}\text{C}$ at the mixing ratio of 1:3, and keeping the temperature at $46 \, ^{\circ}\text{C}$.

After the completion of infection, virus solution was aspirated thoroughly from the dish and 4 ml of the warm X-gal medium containing agarose (previously prepared) was gently added to every dish not to peel off cells. The dish kept open by slightly sliding a lid until the agarose solidified and dried, and thereafter the dish covered, turned upside down, and incubated at 27 °C for 6 days. The plaques were observed under a phase difference microscope to find blue plaques which do not form multinucleate cells. Agarose containing blue recombinant plaques were removed with an aspirating pipet and suspended into 1 ml of TMN-FH medium by pipetting many times. The above process which comprises: infection, 6-day incubation, and isolation of virus containing transfer plasmid DNA is called the "plaque method". The plaque method was repeated using 100 µl of viral suspension. After repeating said process three times, there obtained a recombinant virus having a gene encoding HCV glucoprotein free from contamination with that of wild-type strain.

A viral solution of the primary recombinant virus was prepared by aspirating plaques with a Pasteur pipet, and mixing thoroughly with 1 ml of TMN-FH medium. Because the primary viral solution was low in virus density for infection, it required further treatments for concentration. Thus, 100 µl of viral solution was adsorbed onto Sf9 cells grown in a petri dish (6 cm in diameter) to a semi-confluent, and 4 ml of TMN-FH medium was added thereto and incubated three days. The culture supernatant was recovered to yield a recombinant viral solution for infection.

For the production of HCV structural protein , a suspension of Sf9 cells in TMN-FH medium containing 10% FCS (5 x 10^6 cells/10 ml medium) was added into a Petri dish (9 cm, in diameter) and kept 1 hr for adsorption. After the removal of medium, 250 μ l of recombinant viral solution was added to the dish and spread extensively. To the dish was added 10 ml TMN-FH medium containing 10% FCS and incubated at 27 °C for 4 days. The cells expressing recombinant glycoprotein of HCV were harvested by scraping up and suspended into 1,000 ml of phosphate buffered saline.

Thus, HCV-derived glycoprotein was expressed in Sf9 cells transfected with said virus.

Example 27

35

40

Identification of Expression Products as HCAg

Each expression product obtained in Examples 25 and 26 was identified as HCAg because it reacted immunologically with antiserum obtained from HC patients. Identification was conducted by Western blot technique.

E. coli cells transformed with expression plasmid pCZNS5N or pGEX2218 were grown in the presence of IPTG for 3 hr or a overnight in the same manner as described in Example 25.

Recombinant strains were harvested by centrifuging 1,000 μ l of the cultured broth at 6,500 rpm, 10 min. The pellet was dissolved into a sample solution (50 mM Tris-HCl, pH6.8 containing 2% SDS, 5% mercaptoethanol, 10% glycerin, and 0.005% bromophenol blue) for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml. Sf9 cells infected with viruses which had been treated more than 3 times by plaque method were collected by scraping up and suspended into 1,000 ml of PBS and 100 μ l of the suspension was centrifuged at 6,500 rpm, 10 min to pellet the cells. The pellet was dissolved into a sample solution for SDS-polyacrylamide gel electrophoresis to a final volume of 0.2 ml.

The sample solutions were then boiled at 100 °C for 10 min. Ten μ l of the boiled solution was loaded onto 0.1% SDS-15% polyacrylamide g I (70 x 85 x 1 mm) together with a marker protein LMW Kit E (low-molecular weight marker protein, Pharmacia). Electrophoresis was carried out at a constant current of 30 mA for 45 min in Tris buffer (25 mM Tris, pH 8.3, 192 mM glycine, 0.1% SDS) as electrode buffer. Thereafter, DNA was transferred electophoretically to a nitrocellulose filter by superposing the gel onto a filter BA-83 (S & S), impressing a constant current of 120 mA for about 20 min between gel (cathode) and the filter (anode) as conventionally.

The transcribed filter was cut to remove a part containing a marker protein (referred to as marker filter) and that containing the sample (referred to as sampl filter). The former was stain d with 0.1% (w/v) amideblack 10B and the latter immersed into 0.01 M PBS (pH 7.4) containing 5% (w/v) bovine serum albumin (BSA). Serum from a HC patient was diluted 50 times with 0.01 M PBS (pH 7.4) containing 5% (w/v) BSA. To the sample filter was added 10 µl of diluted serum and the filter allowed to stand for 2 hr at room temperatur. Thereafter, the filter was washed with PBS containing 0.1% (W/V) Tween 20 for 20 min (x3)

The sample filter was then reacted with 10 ml of horseradish peroxidase conjugated anti human IgG (Gappel) at 37 °C for 1 hr and washed with PBS containing 0.1% (w/v) Tween 20 for 20 min (x3). The filter was then immersed into peroxidase-color-producing solution (60 mg 4-chloro-1-naphthol, 20 ml methanol, 80 ml PBS, and 20 µl aqueous hydrogen peroxide). The colored filter was washed with distilled water and compared with the marker filter, demonstrating that colored protein expressed by transformants transformed with plasmid pCZNS5N or pGEX2218 had a reasonable molecular weight as an expression product of inserted HCV gene and was identified as HCAg. The expression product from transformants transformed with pGEX2218 was a fused protein consisting of HCV originated polypeptide and GST and thrombin cleaving site wherein the latter two attached at the N-terminus of the former.

Example 28

Preparation of Clone T7N1-25

[1] Preparation of Clone 1925

Clones N19MX24A-1 (prepared in Example 11[1]) and MX25-1 were ligated using overlapping region by PCR. One μI (about 0.5 to 1 μg/μI) of each DNA fragment from clones N19MX24A-1 and MX25-1 (977 and 849 bp, respectively) was added into a reaction mixture containing 10 µl of 10 x PCR buffer (100 mM Tris-HCI (pH8.3), 500 mM KCI, 15 mM MgCl2, 1% gelatin), 8 µI of 2.5 mM 4 dNTPs, 5 µI each of 20 pmol/µI synthetic primers MS122 and MS152, and 76.5 µl of water. After an intimate mixing, the mixture was heated at 95 °Cfor 5 min, then immediately cooled to 0 °C. One minute later, to the mixture was added 0.5 µl of Taq DNA polymerase (7 U/µI, AmpliTaq™ Takara Shuzo), mixed and overlaid with mineral oil. The sample was then subjected to PCR. PCR was conducted by repeating 10 times of reaction cycle, which comprises the following treatments: at 95 °C for 1 min; at 37 °C for 1 min; and at 72 °C for 2 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by an incubation at 97 °C for 2 min. The mixture was immediately cooled to 0 °C, kept at the same temperature for 2 min, mixed with 0.5 µl of Taq DNA polymerase (7 U/µI, AmpliTaq™ Takara Shuzo), and overlaid with mineral oil. The sample was then treated in the same manner as the above by repeating 15 times of reaction cycle, which comprises the following treatments: at 95 °Cfor 1 min; at 55 °C for 1 min; and at 72 °C for 2 min. After the final treatment at 72 °C for 7 min, the resultant reaction solution was treated with phenol/chloroform then precipitated with ethanol. The two DNA fragments were ligated and amplified by PCR. The ligated DNA sample was fractionated on agarose gel electrophoresis and a gel containing about 1000 bp fragment was excised from the gel (Molecular Cloning (1982) Cold Spring Harbor). The resultant DNA fragment was then modified as described in Example 3 and ligated into Smal site of multi-cloning sites of pUC19, cloned and screened as described in Example 3 to obtain plasmid pUC1925. The resultant clone derived from serum of HC patient was designated as clone 1925.

[2] Preparation of Clone T7N119

45

Plasmid pUCN1-1 contains cDNA clone N1-1 at Smal site of pUC19 from 5' to 3', HindIII site of pUC19 and HCV gene. The plasmid pUCN1-1 was digested with HindIII and Ncol completely and the larger fragment pUCN1HN containing the vector function was isolated. Ten ng of said DNA fragment was ligated to the following synthetic DNAs:

MS168: AGCTTACTAGTTAATACGACTCACTATAGGG (31base pairs, SEQ ID NO: 188)

MS169: CTGGCACCCTATAGTGAGTCGTATTAACTAGTA (33base pairs, SEQ ID NO: 189)

MS170: TGCCAGCCCCTGATGGGGGCGACACTCCACCATAGATCACTCC (44base pairs, SEQ ID NO: 190)

MS171: TCACAGGGGAGTGATCTATGGTGGAGTGTCGCCCCCATCAGGGGG(45base pairs, SEQ ID NO: 191)

MS172: CCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAGC(40base pairs, SEQ ID NO: 192)

MS173: CATGGCTAGACGCTTTCTGCGTGAAGACAGTAGTTCC(37base pairs, SEQ ID NO: 193)

The above DNA fragments ar shown from 5' to 3' termini.

DNA fragments except MS168 and MS173 were kinased at 5' t rminus. A 100 pmol of each of 5'-kinased MS169, MS170, MS171 and MS172, and 20 pmol of each of MS168 and MS173 were ligated in the presence of T4 DNA ligase, and the reaction mixture treated with phenol treatment and ethanol pr cipitation, conventionally. A quarter of the precipitated DNA sample was ligated to 10 ng of pUCN1HN to obtain plasmid pUCT7N1 which comprises from 5, to 3', HindIII site, Spel site, promoter sequence derived from T7RNA polymerase, 5' non-translational region of HCV gene, DNA fragment of a gene encoding the N-terminal region of HCV core protein, at the 5' site of clone N1-1. The resultant plasmid pUCT7N1 contains clone T7N1 between HindIII and Smal sites. Clone T7N1N3N10 was prepared in the same manner as that described in Example 4 [2] except that plasmid pUCT7N1 was used instead of pUCN1-1 having clone N1-1.

Clones T7N1N3N10 and N27N19-1 prepared in Example 11 [2] were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme BamHI, T7N1N3N10 and N27N19-1 were cleaved at 3' site of No. 1332 (G) and No. 3 (G), respectively. The ligation was accomplished on the basis of an assumption that plasmids pUCN1N3N10 and pUCN27N19-1 contain each DNA fragment in the same orientation (on the HCV gene, HindIII site of pUC19 located at 5' site). Thus, plasmid pUCN119 was prepared by digesting pUCN27N19-1 with EcoRI and BamHI to isolate a DNA fragment containing the 5' region of clone N27N19-1 (the fragment comprises clone N27N19-1 attached at the 3' terminus by EcoRI-Smal fragment of plasmid pUCN1N3N10, cloning and screening. Plasmid pUCN119 contains the desired clone T7N119 comprising, from 5' to 3', HindIII site, Spel site, promoter sequence derived from T7RNA polymerase, a part of 5' non-translational region of HCV gene, clones N1-1, N3-1, N10-1, N27-3, N19-1 without overlapping.

[3] Preparation of Clone T7N1-25

Clones T7N119 and 1925 prepared in the above [1] were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the both clones. Upon digestion with restriction enzyme Pvul, clone T7N119 was cleaved at 3' site of No. 288 (T) of basse sequence of clone N19-1 in N19 region which is shown by SEQ ID NO 16, and clone 1925 was cleaved at 3' of No.288 (T). The ligation of T7N119 and 1925 clones was accomplished on the basis of an assumption that plasmids pUCT7N119 and pUC1925 contain each DNA fragment in the same orientation. Thus, plasmid pUCTN119 was prepared by digesting pUC1925 with Pvull and EcoRI to isolate a DNA fragment encoding HCV originated gene (said DNA fragment contains at 3' of said cDNA a EcoRI-Smal fragment of plasmid pUC19), exchanging the Pvull-EcoRI fragment containing 3' region of N19 region of plasmid pUCT7N119 with the fragment obtained from plasmid pUCTN1-25, cloning, and screening.

Plasmid pUCT7N1-25 contains the desired clone T7N1-25 comprising clones T7N119 and 1925 ligated at Pvul site without overlapping.

40 Example 29

45

Preparation of Clone T7N1-30U

[1] Preparation of Clone 1530UNot

The clone 1530U prepared in Example 23[5] contains HindIII site adjacent to 3' site of cDNA of HCV. Plasmid pUC1530U was digested completely with HindIII, blunt ended with T4DNA polymerase conventionally. Ten ng of resultant DNA fragment was ligated to an excess amount of EcoRI-NotI-BamHI adapter (x 100 molar, Toyobo) in the presence of T4 DNA ligase, conventionally. After the phenol treatment and ethanol precipitation, the fragment was digested with NotI, ligated, cloned and screened to yield plasmid pUC1530UNot.

[2] Preparation of Clone T7N1-30U

Clones T7N1-25 prepared in Example 28, MX25N15-1 prepared in Example 17 [4], and 1530UNot obtained in the present Example were ligated by taking advantage of unique restriction site which exists in the overlapping regions of the clones. PUCT7N1-25 was digested with Spel and Pstl and about 1 ng of a DNA fragment T7N1-325SP containing the majority of clone T7N1-25 was extracted from gel. Plasmid

pUCMX25N15-1 was digested with Pstl and EcoT221 and about 1 ng of a DNA fragment MX25N15-1PE containing the majority of clone MX25N15-1 was extracted from gel. Plasmid pUC1530UNot was digested with EcoT22I and Notl and about 1 ng of a DNA fragment 1530UEN containing the majority of clone 1530UNot was isolated from gel.

About 200 ng of each of the above fragments T7N1-25SP, MX25N15-1PE, 1530UEN, and 1 ng of SpecI-NotI fragment of λ ZapII (Strategene) were ligated according to the protocol attached to the kit. It was followed by packaging with GIGAPACKII PACKING EXTRACTS, GOLD (Strategnen). All the procedures including ligation, titer check, amplification of λ DNA, isolation and packaging were conducted according to the teaching of protocol attached thereto. The screening of recombinant phage was carried out for the inserted DNA fragment by isolating 20 white plaques, subcloning into plasmid pBBLUESCRIPT SK (-). Among 2 clones of 20 clones subcloned into plasmid pBBLUESCRIPT SK (-) contained a DNA fragment having three sequences of HCV gene between SpecI and NotI site of said plasmid λ ZapII (from 5' SpecI site to 3': clone T7N1-25SP, MX25N15-1PE and 1530UEN). The resultant plasmid was designated as pT7NI-30U.

The plasmid pT7N1-30U contains a clone T7N1-30U comprising three DNA fragments originated from HCV ligated without overlapping SpecI and NotI sites. Base and amino acid sequence of polypeptide encoded by said clone are shown in SEQ ID NO 101.

Example 30

15

20

Large-Scale-Expression of Polypeptides CORE and C+N23

[1] Preparation of clone CN23

A region of clone N23-1 to be expressed was obtained by PCR using as a template, pUCN23-1 having clone N23-1 prepared in Example 16. The following synthetic DNAs were used as primers.

5' primer:

MS165: 5' GCAAGCTTATGCTGCTGTCGCCCGGGCCCATCT3' (SEQ ID NO: 194) 3' primer:

MS166: 5' GCGAATTCAGATCTTCATCATGTGTTGCAGTCGATCAC 3' (SEQ ID NO: 195)

The synthetic DNA was adjusted to 20 pmol/ml before use.

PCR was carried out in the same manner as described in the above according to Saiki's method in a total volume of 100 μl containing 100 ng of plasmid pUCN23-1, as a template, and 2 μl each of 3' and 5' primers. The reaction mixture was heated at 95 °C for 5 min and quenched at 0 °C. One minute later, to the mixture was added 0.5 μl of Taq DNA polymerase (7 U/ml, AmpliTaq TM Takara Shuzo), mixed thoroughly and overlaid with mineral oil. The sample was reacted by repeating 8 cycles of treatments which comprises: at 95 °C for 1 minute; at 55 °C for 1 min; and at 72 °C for 1 min in DNA Thermal Cycler (Parkin Elmer Cetus). It was followed by 17 times of reaction cycles comprising, at 95 °C for 1 minute; at 65 °C for 1 min; and at 72 °C for 1 min. The resultant reaction solution was extracted with phenol/chloroform, and precipitated with ethanol conventionally. The amplified DNA samples were digested with HindIII and EcoRI, and fractionated on acrylamide gel electrophoresis and extracted.

The DNA fragment was then ligated into HindIII and EcoRI sites of cloning vector pUC19, cloned screened to obtain plasmid pUCN23A. The base sequence of clone N23A shows that it comprises a DNA fragment shown by a base sequence from Nos. 1 to 915 of SEQ ID No 50 having additional DNA fragments attached to the both 5'- and 3'-termini. That is, at its 5'-terminus, the following DNA fragment comprises a HindIII restriction site followed by an initiation codon ATG was attached.

5' GCAAGCTTATG 3'

3' CGTTCGAATAC 5' (SEQ ID NO 155)

And at its 3'-terminus, the following DNA fragment comprises two termination codons, BgIII and EcoRI sites from 5' to 3' was attached.

55

- 5' TGATGAAGATCTGAATTCGC 3'
- 3' ACTACTTCTAGACTTAAGCG 5' (SEQ ID NO 156)

Plasmid pCZCORE obtained in Example 6 [1] was digested with SacII and blunt ended with T4DNA polymerase conventionally, which was followed by the digestion with BgIII and subjected to acrylamide gel electrophoresis. From the gel, a DNA fragment pCZCORE/SB containing a vector part of vector pCZ and the N-terminal region of core protein of HCV was extracted.

In the same manner, plasmid pUCN23A was digested with Smal and Bglll completely and subjected to acrylamide gel electrophoresis. From the gel, a DNA fragment N23A/SB containing the sequence of clone N23-1 was extracted, which fragment contains, from 5' terminus, a base sequence from No.107 (G) to No. 915 (A) of SEQ ID NO 50 and two stop codons and a Bglll site.

Ten ng of a DNA fragment pCZCORE/SB and 100 ng of N23A/SB were ligated conventionally to obtain plasmid pCZCN23, which contains clone CN23 of SEQ ID NO 102between HindIII and BgIII sites.

[2] Modification of expression vector

10

15

20

The improvement of the efficiency of expression was accomplished by making the expression unit in expression vector multiple. Thus, plasmids pCZCORE and pCZCN23 were digested thoroughly with restriction enzymes BamHI and BgIII, and the resultant DNA fragments CORE/BB and CN23/BB encoding a polypeptide derived from HCV was recovered.

The DNA fragment CORE/BB (100 ng) was ligated by T4DNA ligase at 12 °C for 30 minutes according to a conventional method. The resultant material was worked up with phenol treatment and ethanol precipitation, digested with restriction enzymes BamHI and BgIII, digested thoroughly with BgIII, and ligated to plasmid pCZCORE (10 ng) previously dephosphorylated with alkali phosphatase by a conventional method to obtain plasmid pCZCORE tandem 2, 3, 4, 8, 16, in which 2, 3, 4, 6, 12 expression units of polypeptide CORE between BamHI and BgIII sites of plasmid pCZCORE are ligated forwardly in tandem. The same procedure was conducted with the DNA fragment CN23/BB and plasmid pCACN23 to obtain plasmid pCZCN23 tandem 2, 3, 4, 6.

[3] Direct Expression of polypeptides CORE and CN23 in Large Scale

Expression of polypeptide CORE and CN23 in E. coli was conducted using each of expression vector obtained in the above [2] in the same method in Example 6 [1].

For this purpose, conditions such as the timing for induction, species or strains of host cells, number of tandem and the temperature of the culture in the system transformed with pCZCORE were studied.

For example, hosts derived from K12 strain such as JM109, DH5, KS476 and hosts derived from B strain were studied. The degree of expression varies depending on the host. The host derived from B strain and KS476 gave an excellent expression, and the expression amount per culture medium was about 8 to 10 times larger than that obtained using DH5, as host cells. The quantities also varied depending on the time for induction. Thus, 0.5 ml of overnight culture containing transformants (OD 600 = about 1.5) was inoculated into 10 ml bactopepton medium (Difco; 20 g/l bactpepton, 0.2% v/v glycerin, 0.1 M MgSO4, 10 g/l NaCl, 160 μl/l of 0.1% thiamin chloride, 100 mg/l ampicillin) in 10 ml L-shaped tube and cultured at 30°C. IPTG was added either of the time when the conductivity (OD 600) reached to about 0.5, 0.8, 1.2, 2.0 and 3.0 for induction. The cultured broth which was induced when the OD 600 reached to about 0.5 gave the best expression and the amounts of the expression product was highest. The expression was not directly proportional to the number of tandem. For example, when cells transformed with expression plasmid containing in tandem three units of an expression unit CORE/BB, the expression efficiency was low, whereas, it was drastically increased when the plasmid contains 4 units in tandem and kept increase until the number of units becomes 8. However, significant improvement was no more observed and the expression amount was almost the same between cultures containing host cells transformed with tandem 8 and 16. The above studies provided the condition for large-scale-expression of polypeptide CORE as follows. A host cell derived from B strain or KS476 was transformed with pCZCORE tandem 8 and cultured 30°C overnight, inducing the expression when the density reached to about 0.5 (OD 600). Among the plasmid pCZCN23 tandem 2, 3, 4, 6 prepared for the expression of polypeptide CN23, tandem 6 was used and the expression was carried out under the same condition as that used for pCZCORE tandem.

	SEQ ID NO:1	
	SEQUENCE LENGTH: 483 base pairs	
	SEQUENCE TYPE: nucleic acid	
5	STRANDEDNESS: double	
	TOPOLOGY: linear	
	ANTI-SENSE: No	
	MOLECULE TYPE: cDNA to genomic RNA	
10	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: N1-1	
15		
	CTCCACCATA GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC	60
	ATGGCGTTAG TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAGT	120
20	GGTCTGCGGA ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCT TTCTTGGATC	180
20	AACCCGCTCA ATGCCTGGAG ATTTGGGCGT GCCCCCGCGA GACTGCTAGC CGAGTAGTGT	240
	TGGGTCGCGA AAGGCCTTGT GGTACTGCCT GATAGGGTGC TTGCGAGTGC CCCGGGAGGT	300
	CTCGTAGACC GTGCATC ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA ACC	350
25	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr	
	1 5 10	
	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT	398
	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	
30	15 20 25	
	GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG	446
	Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg	
	30 35 40	
35	TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG T	483
	Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg	
	SEQ ID NO:2	
40	SEQUENCE LENGTH: 187 base pairs	
	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
45	ANTI-SENSE: NO	
	MOLECULE TYPE: cDNA to genomic RNA	
	OPICINAL COMMON	

	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: N2-1	
5		
	AGGTCTCGTA GACCGTGCAT C ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA 51	
	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys	
	1 5 10	
10	ACC AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC 99	
	Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly	
	15 20 25	
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC 147	
15	Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro	
	30 35 40	
	AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG T	
	Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg	
20	45 50 55	
	SEQ ID NO:3	
	SEQUENCE LENGTH: 531 base pairs	
25	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA to genomic RNA	
30	ANTI-SENSE: NO	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
05	CLONE: N3-1	
35	CLONE. NJ-1	
	AGGTCTCGTA GACCGTGCAT C ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA ACC 54	
	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr	
40	1 5 . 10	
	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT 102	
	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	
	15 20 25	
45	GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG 150	
	Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg	

			30			•		35					40				
	TTG	GGT	GTG	CGC	GCG	ACT	AGG	AAG	ACT	TCC	GAG	CGG	TCG	CAA	CCT	CGT	198
5	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	
5		45					50					55					
	GGA	AGG	CGA	CAA	CCT	ATC	CCC	AAG	GCT	CGC	CAA	ccc	GAG	GGC	AGG	GCC	246
	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Gln	Pro	Glu	Gly	Arg	Ala	
10	60 [°]					65					70					75 ,	
10	TGG	GCT	CAG	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TTG	294
	Trp	Ala	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	
					80					85					90		
15	GGG	TGG	GCA	GGA	TGG	CTC	CTG	TCA	CCC	CGC	GGC	TCC	CGG	CCT	agt	TGG	342
15	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	
				95					100					105			
	GGC	CCC	ACG	GAC	CCC	CGG	CGT	AGG	TCG	CGT	AAT	TTG	GGT	AAG	GTC	ATC	390
20	Gly	Pro		Asp	Pro	Arg	Arg	Arg	Ser	Arg	Asn	Leu	_	Lys	Val	Ile	
20			110					115					120				
															CGC		438
	Asp		Leu	Thr	Cys	Gly		Ala	qaA	Leu	Met		Tyr	Ile	Pro	Leu	
25		125					130					135					
20															GGC		486
		Gly	Ala	Pro	Leu	_	Gly	Ala	Ala	Arg		Leu	Ala	His	Gly		
	140					145					150					155	
30						GGC											531
	Arg	VAI	Leu	GIU	-	Gly	vaı	Asn	туг		Thr	GIĀ	Asn	Leu			
					160					165					170		
	SEQ	ID I	NO:4														
35	SEQ	UENC	E LEI	NGTH	: 75	5 bas	зе ра	airs									
	SEQ	UENC	E TY	PE: 1	nucl	eic a	acid										

SEQUENCE LENGTH: 755 base pairs SEQUENCE TYPE: nucleic acid STRANDEDNESS: double TOPOLOGY: linear

40 MOLECULE TYPE: cDNA to genomic RNA

ANTI-SENSE: No ORIGINAL SOURCE

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE

CLONE: N10-1

50

45

															CT A		
	V		en T	yr A	la T	уr G	ly A	sn L	eu P	ro G	ly C	ys S	er E	he S	er I	le P	he
5		1				5					10					15	
-															TAC		
	Leu	Leu	Ala		Leu	Ser	Сув	Leu	Thr	Ile	Pro	Ala	Ser	: Ala	Tyr	Gln	•
				20					25					30			
10															TCC		
	Val	Arg		Ala	Ser	Gly	Val		His	Val	Thr	Asn		_	Ser	Asn	
	ma »	3.CM	35					40					45				
															CCC		
15	Ser		114	Val	Thr	GIU		Ala	Asp	Val	Ile			Thr	Pro	Gly	
	WCC.	50	ccc	maa	CMC	999	55					60					
															GTA		
	65	Vul	FIO	Cys	AGT	70	GIU	ABR	ABN	Ser		Arg	Cys	Trp	Val		
20		ACT	CCC	ACG	المليات		ccc	ACC.	220	200	75	»mc	000	3.00	ACG	80	
															Thr		
					85		2224	n.y	ABII	90	Ser	110	PLO	THE	95	THE	
	ATA	CGG	CGT	CAT		GAC	TTG	CTC	CTT		CCA	CCT	CTC	CTC	TGT	ሞሮሮ	337
25															Сув		
		_	_	100		•			105	2				110	0,10	502	
	GCT	ATG	TAT	GTG	GGG	GAT	TTT	TGC	GGA	TCT	GTT	TTC	CTC	_	TCC	CAG	385
															Ser		
30			115					120					125				
	CTG	TTC	ACT	TTC	TCA	CCT	CGC	CGG	TAT	GAG	ACG	GTG	CAA	GAC	TGC	AAT	433
	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	Tyr	Glu	Thr	Val	Gln	Asp	Cys	Asn	
		130					135					140					•
35	TGC	TCA	ATC	TAT	CCC	GGC	CAT	GTA	TCA	GGC	CAT	CGC	ATG	GCT	TGG	GAT	481
	Сув	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg	Met	Ala	Trp	Asp	
	145					150					155				•	160	
															CAG		529
40	Met	Ile	Met	Asn	Trp	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ser	Gln	Leu	
					165					170					175		
															CAC		577
	Leu	Arg	Ile		Gln	Ala	Val	Val		Met	Val	Ala	Gly	Ala	His	Trp	
45				180					185					190			
	GGA	GTC	CTG	GCG	GGC	CTT	GCC	TAC	TAT	TCC	ATG	GTG	GGG	AAC	TGG	GCT	625

	Gly	Val	Leu 195	Ala	Gly	Leu	Ala	Tyr 200	Tyr	Ser	Met	Val		Asn	Trp	Ala	
	220	CITIC		cmm	CITIC	» mc	oma		mma				205				
			TTG														673
5	гур	210	Leu	Val	Vai	Mer	215	ren	Pne	ALA	GIY		Авр	GIY	GTĀ	Thr	
	CAC	_	ACA	ccc	CCN	880		ccc	mac	3.00	200	220	200				
																	723
	225	AGI	Thr	GIY	GIY	230	Val	ATG	TYL	THE		GIN	Ser	Pne	Inr		
10		delete	TCA	CCA	ccc		mem.	CAC	202	3.000	235					240	
			Ser									C					
	rne	rne	ger	ALY	245	PLO	Ser	GIII	ALG	250	GIN						
15				•	•												
		ID I															
			E LEI					_	В		•						
	-		E TY				acid										
20			DNES:			8							·				
		-	Y: 1:														
			E TY		CUNA	to (geno	mic 1	RNA								
			NSE:														
25			L SOI			a	•										
			M: He	-				a-i							,		
			re ex		LMEN.	TAL 3	SOUR	CE									
	CHO	ME: 1	N3N1(J													
30	AGG!	rctc(STA (BACC	STGC	AT C	ATG	AGC	ACA	AAT	CCA	AAA	CCC	CAA	AGA	ÄAA	ACC 54
																Lys	
							1				5	-1-			9	10	
	AAA	CGT	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	_	AAG	TTC	CCG	GGC		102
35			Asn														
	-			15		•	3		20			-1-		25	 1	0-1	
	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	AGG		CCC	AGG	150
			Ile														
40	_		30		_	-		35					40	2		3	
	TTG	GGT	GTG	CGC	GCG	ACT	AGG	AAG	ACT	TCC	GAG	CGG	TCG	CAA	CCT	CGT	198
	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Glu	Pro	Arg	
		45					50					55					
45	GGA	AGG	CGA	CAA	CCT	ATC	CCC	AAG	GCT	CGC	CAA	CCC	GAG	GGC	AGG	GCC	246
50																	

	Gly	Are	g Ar	g Glı	n Pro	Ile	Pro	Ly	s Ala	a Arc	g Gl	n Pro	G1:	u Gl	v Ar	g Ala	
	60)				65					70	_		- .		75	
5	TGG	GC:	CAC	CCC	GGG	TAC	: cc	r TG	cco	CTO	C TAS	r GG(AA :	T GA	G GG(TTG	29/
J	Tr	Ala	a Glr	Pro	Gly	Tyr	Pro	Tr	Pro	Lei	1 Ту	r Gly	, As	n Gl	u Gl	y Leu	
					80)				85	5				90)	
	GGG	TGC	GC#	GG#	1 TGG	CTC	CTG	TC	A CCC	CGC	GGG	TCC	CGG	G CC	r ag	TGG	342
10	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Se:	Pro	Arg	Gly	, Sei	Arg	g Pro	Sei	Trp	
10				95	i				100)				10!	5		
	GGC	CCC	ACG	GAC	ccc	CGG	CGI	AGG	TCG	CGI	CAA !	TTG	GG	C AAC	GTC	ATC	390
	Gly	Pro	Thr	Asp	Pro	Arg	Arg	Arg	Ser	Arg	Asr	Leu	Gl	, Lys	Val	Ile	
15			110					115	3				120)			
15	GAT	ACC	CTC	ACA	TGC	GGC	TTC	GCC	GAT	CTC	ATG	GGT	AC	TTC	CGC	TCG	438
	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Авр	Leu	Met	Gly	Туг	: Ile	Pro	Leu	
	•	125					130			•	•	135					
20	GTC	GGC	GCC	ccc	CTA	GGG	GGC	GCT	GCC	AGG	GCT	CTA	GCG	CAT	GGC	GTC	486
20	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	Gly	Val	
	140					145					150					155	
	CGG	GTT	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	CTG	CCT	GGT	534
05	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn	Leu	Pro	Gly	
25					160					165					170		
	TGC	TCC	TTT	TCT	ATC	TTC	CTT	TTG	GCT	TTG	CTG	TCC	TGT	TTG	ACC	ATC	582
	Суѕ	Ser	Phe		Ile	Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu	Thr	Ile	
••				175					180					185			
30	CCA	GCT	TCC	GCC	TAC	CAA	GTG	CGC	AAC	GCG	TCC	GGG	GTG	TAC	CAT	GTC	630
	Pro	Ala	Ser	Ala	Tyr	Gln	Val	Arg	Asn	Ala	Ser	Gly	Val	Tyr	His	Val	
			190					195					200				
	ACG	AAC	GAC	TGC	TCC	AAC	TCA	AGT	ATT	GTG	TAT	GAG	GCG	GCG	GAC	GTG	678
35	THE	ASN	Asp	Cys	Ser	Asn		Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Val	
	A mm	205	~~~				210					215					
	AIT	ATG	CAC	ACC	ccc	GGG	TGC	GTG	CCC	TGC	GTC	CGG	GAG	AAC	AAT	TCC	726
	116	Met	HIS	Tyr	Pro		Сув	Val	Pro	Сув	Val	Arg	Glu	Asn	Asn	Ser	
40	220					225					230					235	
	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACT	CCC	ACG	CTT	GCG	GCC	AGG	AAC	AGC	774
	ser	Arg	Cys	Trp		Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	
	100	3.00			240					245					250		
45	AGC .	ATC	CCC	ACT	ACG .	ACA .	ATA	CGG	CGT	CAT	GTC	GAC	TTG	CTC	GTT	GGG	822
	per	тте	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	Gly	

				255					260					265			
	GCA	GCT	GCT	CTC	TGT	TCC	GCT	ATG	TAT	GTG	GGG	GAT	TTT	TGC	GGA	TCT	870
5	Ala	Ala	Ala	Leu	Сув	Ser	Ala	Met	Tyr	Val	Gly	Asp	Phe	Сув	Gly	Ser	
Ū			270					275					280				
						CAG											918
	Val		Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	Tyr	Glu	
10		285					290					295					
						AAT											966
		Val	Gln	Asp	Сув	Asn	Сув	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	
	300					305					310					315	
15						GAT											1014
	HIS	Arg	met	ATG		qaA	Met	ITe	Met		Trp	Ser	Pro	Thr		Ala	
	CTL	CITC	CER	mac	320	Om s	OMO.	000	> == 0	325					330		
						CTA Leu											1062
20	200	***	vul	335	GIII	Tea	nea	ALG	340	PIO	GIN	ATG	Val	345	Авр	Met	
	GTG	GCG	GGG		CAC	TGG	GGA	GTC		GCG	ccc	Cutur	GCC		ጥልጥ	TCC .	1110
						Trp											1110
		•	350			_	•	355			1		360	-1-	-1-		
25	ATG	GTG	GGG	AAC	TGG	GCT	AAG	GTC	TTG	GTT	GTG	ATG	CTG	CTC	TTC	GCC	1158
	Met																
		365					370					375					
	G GT	GTT	GAC	GGG	GGG	ACC	CAC	GTG	ACA	GGG	GGA	AAG	GTA	GCC	TAC	ACC	1206
30	Gly	Val	Asp	Gly	Gly	Thr	His	Val	Thr	Gly	Gly	Lys	Val	Ala	Tyr	Thr	
	380					385					390					395	
	ACC																1254
	Thr	Gln	Ser	Phe		Ser	Phe	Phe	Ser	Arg	Gly	Pro	Ser	Gln	Arg	Ile	
35					400					405					410		
	CAGC																1258
	Gln																
40	SEQ	TD K	i0 • 6														
	SEQU			ican .	. 156	: 4 h-		. a d	_								
	SEQU							all'8	•								
	STRA						CIU										
45	TOPO					•											
	MOLE				-	to o	enon	ic =	AN	٠							
						9											

	ANTI-SENSE: No
	ORIGINAL SOURCE
5	ORGANISM: Hepatitis C virus
3	IMMEDIATE EXPERIMENTAL SOURCE
	CLONE: N1N3N10

10	CTCC	ACC	ATA (GATC.	ACTC	CC C	TGTG	AGGA	A CT	ACTG	TCTT	CAC	GCAG	AAA	GCGT	CTAGCC	60
70	ATGG	CGT	rag :	ratg.	agtg'	TC G	TGCA	GCCT	C CA	GGAC	cccc	CCI	CCCG	GGA	GAGC	CATAGT	120
	GGTC	TGC	GGA 1	ACCG	GTGA(GT A	CACC	GGAA	T TG	CCAG	GACG	ACC	GGGT	CCT	TTCT	TGGATC	180
	AACC	CGC	rca i	ATGC	CTGG	AG A	TTTG	GGCG	r GC	CCCC	GCGA	GAC	TGCT	AGC	CGAG	Tagtgt	240
45	TGGG	TCG	CGA A	AAGG	CCTT	gt G	GTAC	TGCC	T GA	TAGG	GTGC	TTG	CGAG	TGC	CCCG	GGAGGT	300
15	CTCG	TAG	ACC (STGC.	ATC 2	ATG .	AGC .	ACA .	AAT (CCA .	AAA	CCC	CAA	AGA	AAA .	ACC	350
					1	Met :	Ser	Thr .	Asn :	Pro :	Lys	Pro	Gln .	Arg	Lys	Thr	
						1				5					10		
	AAA	CGT	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	GGT	398
20	Lys	Arg	Asn	Thr	Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	
				15					20					25			
	GGT																446
or	Gly	Gln		Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	
25			30					35					40				
	TTG																494
	Leu		Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	
•		45					50					55					
30	GGA																542
		Arg	Arg	Gln	Pro		Pro	Lys	Ala	Arg	Gln	Pro	Glu	Gly	Arg	Ala	
	60					65					70					75	
	TGG																590
35	Trp	Ala	Gln	Pro		Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	
					80					85					90		
	GGG																638
	Gly	Trp	Ala		Trp	Leu	Leu	Ser		Arg	Gly	Ser	Arg	Pro	Ser	Trp	
40				95					100					105			
	GGC																686
	Gly	Pro		Asp	Pro	Arg	Arg		Ser	Arg	Asn	Leu		Lys	Val	Ile	
			110					115					120				
45	GAT																734
	Asp	Thr	Leu	Thr	Сув	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu	

		125					130					135					
	GTC	GGC	GCC	ccc	CTA	GGG	GGC	GCT	GCC	AGG	GCT		GCG	СУТ	CCC	CTC	782
5						Gly											702
3	140					145				3	150		1114	1110	GIY	155	
	CGG	GTT	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA		GGG	ААТ	CTC	ር ር		830
						Gly											030
10	•				160	-			-4-	165		1		204	170	GIY	
10	TGC	TCC	TTT	TCT	ATC	TTC	CTT	TTG	GCT		CTG	TCC	TGT	THE STEEL		ልጥሮ	878
						Phe											070
				175					180				-4-	185			•
15	CCA	GCT	TCC	GCC	TAC	CAA	GTG	CGC	AAC	GCG	TCC	GGG	GTG		САТ	GTC	926
75						Gln											320
			190					195				-	200	•			
	ACG	AAC	GAC	TGC	TCC	AAC	TCA	AGT	ATT	GTG	TAT	GAG	GCG	GCG	GAC	GTG	974
20						Asn											
20	•	205					210					215			_		
	ATT	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	GTC	CGG	GAG	AAC	AAT	TCC	1022
	Ile	Met	His	Thr	Pro	Gly	Cys	Val	Pro	Сув	Val	Arg	Glu	Asn	Asn	Ser	
25	220					225					230					235	
						GCG											1070
	Ser	Arg	Сув	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	
					240					245					250		
30						ACA											1118
00	Ser	Ile	Pro		Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	Gly	
				255					260					265			
	GCA	GCT	GCT	CTC	TGT	TCC	GCT	ATG	TAT	GTG	GGG	GAT	TTT	TGC.	GGA	TCT	1166
35	Ala	Ala		Leu	Сув	Ser	Ala	Met	Tyr	Val	Gly	Asp	Phe	Сув	Gly	Ser	
	~~~		270			_		275					280				
						CAG											1214
	vai		Leu	Val	Ser	Gln		Phe	Thr	Phe	Ser		Arg	Arg	Tyr	Glu	
40	3.00	285	<b>~</b> >>	a. a			290					295					
	Mb~	G1G	CAA	GAC	TGC	AAT	TGC	TCA	ATC	TAT -	ccc	GGC	CAT	GŤA	TCA	GGC	1262
	300	vaı	GIN	Asp	Cys	Asn	Cys	Ser	Ile	Tyr		Gly	His	Val	Ser	Gly	
		ccc	n m∽	CCM	maa	305	<b>.</b>				310					315	
45						GAT											1310
	HTO	мц	net	urg		Asp	Tem	тте			Trp	Ser	Pro	Thr		Ala	
					320					325					330		

	CTA	GTG	GTA	TCG	CAG	CTA	CTC	CGG	ATC	CCA	CAA	GCC	GTC	GTG	GAT	ATG	1358
	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro	Gln	Ala	Val	Val	Asp	Met	
5				335					340					345			
•	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	TAC	TAT	TCC	1406
	Val	Ala	Gly	Ala	His	Trp	Gly	Val	Leu	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	
			350					355					360				
10	ATG	GTG	GGG	AAC	TGG	GCT	AAG	GTC	TTG	GTT	GTG	ATG	CTG	CTC	TTC	GCC	1454
	Met	Val	Gly	Asn	Trp	Ala	Lys	Val	Leu	Val	Val	Met	Leu	Leu	Phe	Ala	
		365					370					375					
	GGT	GTT	GAC	GGG	GGG	ACC	CAC	GTG	ACA	GGG	GGA	AAG	GTA	GCC	TAC	ACC	1502
15	Gly	Val	Asp	Gly	Gly	Thr	His	Val	Thr	Gly	Gly	Lys	Val	Ala	Tyr	Thr	
	380					385					390					395	
	ACC	CAG	AGC	TTT	ACA	TCC	TTC	TTT	TCA	CGA	GGG	CCG	TCT	CAG	AGA	ATC	1550
	Thr	Gln	Ser	Phe	Thr	Ser	Phe	Phe	Ser	Arg	Gly	Pro	Ser	Gln	Arg	Ile	
20		_			400					405					410		
	CAG	С															1554
	Gln																
	<b>G</b> 220																
25	SEQ																
					370		_	irs									
					ucle uble		cid										
	TOPO					,											
30						<b>.</b>											
	MOLE				DNA	to g	enom	IC R	.NA								
	ORIG																
	ORGA				+10	c:											
35	IMME							72									
	CLON			I DILI	TIBLE I	ALI S	OORC	E.									
	GCAA	GCTT	ATC	AGC	ארא	ከልሞ	CCX	***	000	<i>a</i>							
40																•	47
			1		Thr	11011	5	пур	PLO	GIN	Arg		Thr	rys	Arg		
	AAC I	ACC :	_		CGC 4	CCA 4	_	CAC :	ርጥር ነ		mmc 4	10	700	70m -		77.6	0
	Asn (	Thr .	Asn i	Arg	Ara I	Pro (	Gln	Agn '	Ual 1	nag :	Dho 1	Des i	3GC (	36T (	3GT (	JAG 33	95
<b>1</b> 5		15		9	- <b>5</b> -		20	E	-u <u>.</u>	-Ja 1	. 116	25	эт.У. (	arā (	эт.Х (	эΤIJ	
	ATC (		GGT (	GGA (	GTT :	rac (	_	י בארים	CCG 4	יפר י	NGC: 4		יייי י	NCC 1	nano a	7 <i>0</i> m	142
					'					.GC 1		-GC (	-CC 1	. ၁၀	iTG (	3G.T.	143

	Ile 30	Val	Gly	Gly	Val	<b>Tyr</b> 35	Leu	Leu	Pro	Arg	Arg 40	Gly	Pro	Arg	Leu	Gly 45	
	GTG	CGC	GCG	ACT	AGG	AAG	ACT	TCC	GAG	CGG	CCG	CAA	CCT	CGT	GGA	AGG	191
5	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Pro	Gln	Pro	Arg	Gly	Arg	
					50					55					60		
	CGA	CAA	CCT	ATC	CCC	AAG	GCT	CGC	CAA	CCC	GAG	GGT	AGG	GCC	TGG	GCT	239
	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Gln	Pro	Glu	Gly	Arg	Ala	Trp	Ala	
10				65					70					75			
						TGG											287
	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	Gly	Trp	
15			80					85					90				
15						TCA											335
	Ala		Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	
		95					100					105					
20						AGG	TGA	AGAT(	CTG 1	AATT(	CGC						370
		Asp	Pro	Arg	Arg												
	110					115					•						
	CBO	TD 1	70 - O														
25		ID 1		acmu.	. 12	64 h.			_								
						64 ba eic a		_	3								
	_		ONES				icia										
			7: 1:			=											
30					_	to q	7070	mic I	DNA								
			NSE:		JUMA	ro i	Jenor	uic i	WA					٠			
			L SOT														
				_	itia	C vi	i ma										
35				_		FAL S		CE									
			HN3N:					-									
40	GCA	AGCT"	r ato	G AG	C AC	A AA	r cc	A AA	A CC	C CA	A AG	A AA	A AC	C AA	A CGT	r	47
40						r Ası											•
				L				5				10			••	•	
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	GGT	GGT	CAG	95
45						Pro											
-		15		_	_		20	-		•		25	-	-	•		
	ATC	GTT	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	AGG		ccc	AGG	TTG	GGT	143

	Ile	Val	Glv	Glv	Val	Tvr	Leu	Leu	Pro	Ara	Ara	Glw	Pro	). Ara	T.au	Gly	
	30		2			35				9	40	017	110	ALG	Dea	45	
	GTG	CGC	GCG	ACT	AGG	AAG	ACT	TCC	GAG	CGG		CAA	CCT	CGT	GGA	AGG	191
5							Thr										
					50					55				_	60		
	CGA	CAA	CCT	ATC	CCC	AAG	GCT	CGC	CAA	CCC	GAG	GGC	AGG	GCC	TGG	GCT	239
	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Gln	Pro	Glu	Gly	Arg	Ala	Trp	Ala	
10				65		*			70					75	_		
	CAG	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TTG	GGG	TGG	287
	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	Gly	Trp	
15			80					85					90				
							CCC										335
	Ala		Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	
		95					100					105					
20							TCG										383
		Авр	Pro	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	
	110					115					120					125	
							GAT										431
25	Leu	Thr	Суѕ	Gly		Ala	Asp	Leu	Met		Tyr	Ile	Pro	Leu	Val	Gly	
					130					135					140		
							GCC										479
	AIA	Pro	Leu		GIĀ	Ala	Ala	Arg		Leu	Ala	His	Gly		Arg	Val	
30	CITIC	CNC	<b>G2.0</b>	145	ama				150					155			
							TAT										527
	neu	GIU	160	GTĀ	VAI	Asn	Tyr	_	Thr	GLY	Asn	Leu		Gly	Cys	Ser	
35	بالعلمان	ጥርጥ		ጥጥር	Cultu	ጥጥረ	GCT	165	CTC	mcc	m~m	mmc	170	3.000	663	cam	c = c
00							Ala										575
		175					180		Dou	Del	Cyb	185	****	110	FIO	VIG	
	TCC		TAC	CAA	GTG	CGC	AAC	GCG	TCC	GGG	GTG	_	ሮልሞ	GTC	ACG	AAC	623
40							Asn										023
	190		-			195				1	200	-1-		***	****	205	
	GAC	TGC	TCC	AAC	TCA		ATT	GTG	TAT	GAG		GCG	GAC	GTG	ልሞጥ		671
							Ile										0.1
45	_				210				-	215				<b>-</b>	220		
	CAC	ACC	CCC	GGG	TGC	GTG	ccc	TGC	GTC		GAG	AAC	ААТ	TCC		CGC	719
							Pro										3
				_	_			-		-							

				225					230	1				235			
	TGC	TGG	GTA	GCG	CTC	ACT	ccc	ACG	CTT	GCG	GCC	AGG	AAC	AGC	AGC	ATC	767
_	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	Ser	Ile	
5			240					245					250				
	CCC	ACT	ACG	ACA	ATA	CGG	CGT	CAT	GTC	GAC	TTG	CTC	GTT	GGG	GCA	GCT	815
	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	Gly	Ala	Ala	
10		255					260					265		-			
10	GCT	CTC	TGT	TCC	GCT	ATG	TAT	GTG	GGG	GAT	TTT	TGC	GGA	TCT	GTT	TTC	963
	Ala	Leu	Сув	Ser	Ala	Met	Tyr	Val	Gly	Asp	Phe	Сув	Gly	Ser	Val	Phe	
	270					275					280					285	
15	CTC	GTC	TCC	CAG	CTG	TTC	ACT	TTC	TCA	CCT	CGC	CGG	TAT	GAG	ACG	GTG	911
	Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	Tyr	Glu	Thr	Val	
					290					295					300		
	CAA	GAC	TGC	AAT	TGC	TCA	ATC	TAT	CCC	GGC	CAT	GTA	TCA	GGC	CAT	CGC	959
20	Gln	Asp	Сув	Asn	Сув	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg	
				305					310					315			
	ATG	GCT	TGG	GAT	ATG	ATA	ATG	AAT	TGG	TCA	CCT	ACA	ACA	GCC	CTA	GTG	1007
	Met	Ala	Trp	Asp	Met	Ile	Met	Asn	Trp	Ser	Pro	Thr	Thr	Ala	Leu	Val	
25			320					325					330				
	GTA	TCG	CAG	CTA	CTC	CGG	ATA	CCA	CAA	GCC	GTC	GTG	GAT	ATG	GTG	GCG	1015
	Val		Gln	Leu	Leu	Arg	Ile	Pro	Gln	Ala	Val	Val	Asp	Met	Val	Ala	
		335					340					345					
30	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	TAC	TAT	TCC	ATG	GTG	1103
	Gly	Ala	His	Trp	Gly	Val	Leu	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	
	350					355					360					365	
												CTC					1151
35	Gly	Asn	Trp	Ala		Val	Leu	Val	Val	Met	Leu	Leu	Phe	Ala	Gly	Val	
					370					375					380		
	GAC	GGG	GGG	ACC	CAC	GTG	ACA	GGG	GGA	AAG	GTA	GCC	TAC	ACC .	ACC	CAG	1199
40	Asp	Gly	Gly	Thr	His	Val	Thr	Gly	Gly	Lys	Val	Ala	Tyr	Thr	Thr	Gln	
40				385					390					395			•
												CAG					1247
	Ser	Phe		Ser	Phe	Phe	Ser	Arg	Gly	Pro	Ser	Gln .	Arg	Ile			
45	<b></b> -		400					405					410				
	TGAA	GATC	TG A	ATTC	GC												1264

SEQ ID NO:9

55

	SEQUENCE LENGTH: 483 base pairs	
	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
5	TOPOLOGY: linear	
	ANTI-SENSE: No	
	MOLECULE TYPE: cDNA to genomic RNA	
	ORIGINAL SOURCE	
10	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: N1-2	
	$\cdot$	
15	CTCCACCATA GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC	60
	ATGGCGTTAG TATGAGTGTC GTGCAGCCTC CAGGCCCCCC CCTCCCGGGA GAGCCATAGT	120
	GGTCTGCGGA ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCT TTCTTGGATC	180
	AACCCGCTCA ATGCCTGGAG ATTTGGGCGT GCCCCCGCGA GACTGCTAGC CGAGTAGTGT	240
20	TGGGTCGCGA AAGGCCTTGT GGTACTGCCT GATAGGGTGC TTGCGAGTGC CCCGGGAGGT	300
	CTCGTAGACC GTGCATC ATG AGC ACA AAT CCT AAA CCC CAA AGA CAA ACC	350
	Met Ser Thr Asn Pro Lys Pro Gln Arg Gln Thr	
25	1 5 10	
20	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT	398
	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	
	15 20 25	
30	GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG	446
30 .	Gly Gln Fle Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg	-
	30 35 40	
	TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG T	483
35	Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg	
•	45 50	
	SEQ ID NO:10	
	SEQUENCE LENGTH: 483 base pairs	
40	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
	ANTI-SENSE: NO	
45	MOLECULE TYPE: cDNA to genomic RNA	
	ORIGINAL SOURCE	

	ORGANISM: nepatitis C Virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: S1-1	
5		
	CTCCACCATA GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC	60
	ATGGCGTTAG TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAGT	120
	GGTCTGCGGA ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCT TTCTTGGATT	180
10	AACCCGCTCA ATGCCTGGAG ATTTGGGCGT GCCCCCGCGA GACCGCTAGC CGAGTAGTGT	240
	TGGGTCGCGA AAGGCCTTGT GGTACTGCCT GATAGGGTGC TTGCGAGTGC CCCGGGAGGT	300
	CTCGTAGACC GTGCACC ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC	350
45	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr	
15	1 5 10	•
	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT GGT	398
	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	
20	15 20 25	
20	GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG	446
	Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg	
	30 35 40	
25	TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG T	483
	Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg	
	45 50	
30	SEQ ID NO:11	
	SEQUENCE LENGTH: 483 base pairs	
	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
35	TOPOLOGY: linear	
	ANTI-SENSE: NO	
	MOLECULE TYPE: cDNA to genomic RNA	
	ORIGINAL SOURCE	
40	ORGANISM: Hepatitis C virus	•
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: S1-2	
	CMCCO CCO MA CO MCCCC CMCCCC CMCCCO CCO CO C	
45	CTCCACCATA GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC	60
	ATGGCGTTAG TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAGT	120
	- LANCE THE LANCE DESCRIPTION OF DESCRIPTION OF THE PROPERTY O	104

	AACCCGCTCA ATGCCTGGAG ATTTGGGCGT GCCCCCGCGA GACCGCTAGC CGAGTAGTGT	240
	TGGGTCGCGA AAGGCCTTGT GGTACTGCCT GATAGGGTGC TTGCGAGTGC CCCGGGAGGT	300
_	CTCGTAGACC GTGCACC ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC	350
5	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr	
	1 5 10	
	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT GGT	398
	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	
10	15 20 25	
	GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG	446
	Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg	
	30 35 40	
15	TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG T	483
	Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg	
	45 50	
20	SEQ ID NO:12	
	SEQUENCE LENGTH: 483 base pairs	
	SEQUENCE TYPE: nucleic acid	
05	STRANDEDNESS: double	
25	TOPOLOGY: linear	
	ANTI-SENSE: NO	
	MOLECULE TYPE: cDNA to genomic RNA ORIGINAL SOURCE	
30	ORGANISM: Hepatitis C virus	
30	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: S1-3	
	CLOND: 01-3	
35	CTCCACCATA GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC	60
-	ATGGCGTTAG TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAGT	120
	GGTCTGCGGA ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCT TTCTTGGATT	180
	AACCCGCTCA ATGCCTGGAG ATTTGGGCGT GCCCCCGCGA GACCGCTAGC CGAGTAGTGT	240
40	TGGGTCGCGA AAGGCCTTGT GGTACTGCCT GATGGGGTGC TTGCGAGTGC CCCGGGAGGT	300
••	CTCGTAGACC GTGCACC ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC	350
	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr	330
	1 5 10	
45	AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT GGT	398
.5	Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly	570

			15		•			20	0				2	5			
GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC			G CC	G CGC	AGG			C AG	G	446
											Arg						
		30					35				•	40		•		•	
TTG (	GGT	GTG	CGC	GCG	ACI	AGC	S AAC	G ACT	r TC	C GAC	G CGG	T					483
Leu	Gly	Val	Arg	Ala	Thr	Arç	g Lys	Th:	c Se	r Glu	ı Arg	ſ					
•	45					5€	)										
SEQ	ID 1	NO:1	3														
SEQU	ENC	E LE	ngth	: 33	9 ba	se j	pairs	3									
EQU:	ENC	E TY	PE:	nucl	eic	acio	1										
STRA	NDE	DNES	S: d	oubl	e												
OPO	LOG	Y: 1	inea	r													
ANTI	-SE	NSE:	No										•				
-			URCE														
DRGA	NIS		epat														
		BB B	XPER	IMRN	TAL	SOUT	RCE										
	DIA:	LE E															
IMME:	E: .1	N27-	1				rg gr	rc ca	ነጥ ልባ	rc cr	ואב פר	e ee	יב כו		ארי ייא	3C C	2h 40
IMME: CLON:	E: ]	N27- G AT	1 C CC	A CA	A GC	C GT				et Va	rg gc				is T		GA 49 ly
LON: C	E: ] CG0 Arg	N27- G AT g Il	1 C CC e Pr	A CA o Gl	A GC n Al	C GT a Va 5	al Va	al As	вр Ме	et Va	al Al	a Gl	y A.	la H	is T	rp G] 15	
LON:	CGC Arc	N27- G AT g Il	1 C CC e Pr GCG	A CA o Gl GGC	A GC n Al CTT	C GT a Va 5 GCC	al Va	al As	sp Me	et Va 1 ATG	al Al LO	a Gl GGG	y A: Aac	la H	is T	rp G] 15 AAG	ly
LON: C G	CGC Arc	N27- G AT G Il CTA	C CC e Pr GCG   Ala	A CA o Gl GGC Gly 20	A GC n Al CTT Leu	C GT a Va 5 GCC Ala	al Va TAC Tyr	al As TAT Tyr	TCC Ser 25	et Va 1 ATG Met	al Al 10 GTG Val	a Gl GGG Gly	y A. AAC Asn	TGG Trp 30	GCT Ala	rp Gi 15 AAG Lys	ly
MMEI LON C G: V:	CGC Arc	N27- G AT G Il CTA Leu	C CC e Pr GCG Ala	A CA o Gl GGC Gly 20 GTG	A GO n Al CTT Leu ATG	CTG	TAC TYT	TAT TYF	TCC Ser 25 GCC	ATG Met	al Al lO GTG Val GTT	a Gl GGG Gly GAC	y A AAC Asn GGG	TGG Trp 30	GCT Ala	rp G1 15 AAG Lys CAC	ly
C G G	CGC Arc	N27- G AT G Il CTA Leu	C CC e Pr GCG Ala	A CA o Gl GGC Gly 20 GTG	A GO n Al CTT Leu ATG	CTG	TAC TYT	TAT TYF	TCC Ser 25 GCC	ATG Met	al Al 10 GTG Val	a Gl GGG Gly GAC	y A AAC Asn GGG	TGG Trp 30	GCT Ala	rp G1 15 AAG Lys CAC	ly 97
C G G V	CGC Arc	N27- G AT G Il CTA Leu	C CC e Pr GCG Ala GTT Val	A CA o Gl GGC Gly 20 GTG	A GC n Al CTT Leu ATG Met	C GTG	TAC Tyr CTC Leu	TAT Tyr TTC Phe 40	TCC Ser 25 GCC Ala	ATG Met GGT	al Al 10 GTG Val GTT Val	a Gl GGG Gly GAC Asp	y A AAC Asn GGG Gly 45	TGG Trp 30 AGG Arg	GCT Ala ACC	rp GJ 15 AAG Lys CAC His	ly 97
CLON:  G' V' G' G'	CGC Arg TC Cal 1	N27- G AT g Il CTA Leu Leu ACA	C CC e Pr GCG Ala GTT Val	A CA o G1 GGC G1y 20 GTG Val	A GC n Al CTT Leu ATG Met	CTG Leu GTA	TAC Tyr CTC Leu	TAT Tyr TTC Phe 40 TAC	TCC Ser 25 GCC Ala	ATG Met GGT Gly	ol Al 10 GTG Val GTT Val	a Gl GGG Gly GAC Asp	AAC Asn GGG Gly 45	TGG Trp 30 AGG Arg	GCT Ala ACC Thr	rp G) 15 AAG Lys CAC His	ly 97
MME C C G V	CGC Arg TC Cal 1	N27- G AT G II CTA Leu TTG Leu ACA	C CC e Pr GCG Ala GTT Val	A CA o G1 GGC G1y 20 GTG Val	A GC n Al CTT Leu ATG Met	CTG Leu GTA	TAC Tyr CTC Leu	TAT Tyr TTC Phe 40 TAC	TCC Ser 25 GCC Ala	ATG Met GGT Gly	al Al 10 GTG Val GTT Val	a Gl GGG Gly GAC Asp	AAC Asn GGG Gly 45	TGG Trp 30 AGG Arg	GCT Ala ACC Thr	rp G) 15 AAG Lys CAC His	97 145
G' Va	CGC Arg	N27- G AT g Il CTA Leu TTG Leu TTG ACA 50	C CC e Pr GCG Ala GTT Val 35 GGA Gly	A CA o G1 GGC G1y 20 GTG Val GGG GIY	A GC n Al CTT Leu ATG Met AAG	C GTA.	TAC Tyr CTC Leu GCC Ala 55	TAT Tyr TTC Phe 40 TAC Tyr	TCC Ser 25 GCC Ala	ATG Met GGT Gly ACC	ol Al 10 GTG Val GTT Val CAG Gln	a G1 GGG G1Y GAC Asp AGG Arg	AAC Aan GGG Gly 45 TTT Phe	TGG Trp 30 AGG Arg ACA	GCT Ala ACC Thr TCC Ser	rp Gl 15 AAG Lys CAC His TTC	97 145
MME LON: C G: V: G: V:	CGC Arc TC ( al 1 TTC 1 al 1 TTC 1	N27- G AT G AT ILEU TTG ACA Fhr 50	C CC e Pr GCG Ala GTT Val 35 GGA GIY	A CA o G1 GGC G1y 20 GTG Val GGG G1y	A GC n Al CTT Leu ATG Met AAG Lys	C GTA CTG Leu GTA Val	TAC Tyr CTC Leu GCC Ala 55 CAG	TAT TYT TTC Phe 40 TAC TYT	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr	ol Al IO GTG Val GTT Val CAG Gln	a G1 GGG G1y GAC Asp AGG Arg 60 GTA	AAC ABN GGG Gly 45 TTT Phe	TGG Trp 30 AGG Arg ACA Thr	GCT Ala ACC Thr TCC Ser	rp Gl 15 AAG Lys CAC His TTC Phe	97 145
C C G G V G G V G G T P P	CGC Arc TC ( aal 1 TC ( tal 1) TTG ( tal 1) TTG ( tal 1)	N27- G AT G AT ILEU TTG ACA Fhr 50	C CC e Pr GCG Ala GTT Val 35 GGA GIY	A CA o G1 GGC G1y 20 GTG Val GGG G1y	A GC n Al CTT Leu ATG Met AAG Lys	C GTA CTG Leu GTA Val TCC Ser	TAC Tyr CTC Leu GCC Ala 55 CAG	TAT TYT TTC Phe 40 TAC TYT	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr	ol Al 10 GTG Val GTT Val CAG Gln	a G1 GGG G1y GAC Asp AGG Arg 60 GTA	AAC ABN GGG Gly 45 TTT Phe	TGG Trp 30 AGG Arg ACA Thr	GCT Ala ACC Thr TCC Ser	rp Gl 15 AAG Lys CAC His TTC Phe	97 145 193
CLON: C G G V G V G T T G P P G G G T T G P P G G G T T G P P G G G T T G T T G T T G T T T T	CGGArg	N27- G AT G II CTA Leu TTG SO FCA SO FCA	C CC e Pr GCG Ala GTT Val 35 GGA Gly CGA	A CA o G1 GGC G1y 20 GTG Val GGG G1y GGG	A GC n Al CTT Leu ATG Met AAG Lys CCG Pro	C GTA. CTG Leu  GTA. Val  TCC Ser 70	TAC TYT CTC Leu GCC Ala 55 CAG GIn	TAT Tyr TTC Phe 40 TAC Tyr AAA Lys	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr CAA Gln	GTT Val  CAG GIN  CTT Leu 75	a Gl GGG Gly GAC Asp AGG Arg 60 GTA	AAC Asn GGG Gly 45 TTT Phe AAC Asn	TGG Trp 30 AGG Arg ACA Thr	GCT Ala ACC Thr TCC Ser	TTC Phe GGC Gly 80	97 145 193
IMME CLON C G V G V T	CGC Arc	N27- G AT G AT ILEU TTG ACA Fhr SO TCA TGG	CCC CCC CCC CCC CCC CCC CCC CCC CCC CC	A CA o G1 GGC G1y 20 GTG Val GGG G1y GGG G1y	A GC n Al CTT Leu ATG Met AAG Lys CCG Pro	CTG-Leu-CTAL-Val-TCC-Ser-70	TAC Tyr CTC Leu GCC Ala 55 CAG Gln	TAT TYT  TTC Phe 40 TAC TYT  AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC Ile	ATG Met GGT Gly ACC Thr CAA Gln	GTG Val GTT Val CAG GIn CTT Leu 75	a Gl GGG Gly GAC Asp AGG Arg 60 GTA Val	AAC AS GGG G1y 45 TTT Phe AAC AS GAC	TGG Trp 30 AGG Arg ACA Thr ACT Thr	GCT Ala ACC Thr TCC Ser AAC Asn	TP GI AAG Lys CAC His TTC Phe GGC Gly 80 AAC	97 145 193
CLONIC C	CGC Arc	N27- G AT G AT ILEU TTG ACA Fhr SO TCA TGG	CCC CCC CCC CCC CCC CCC CCC CCC CCC CC	A CA o G1 GGC G1y 20 GTG Val GGG G1y GGG G1y	A GC n Al CTT Leu ATG Met AAG Lys CCG Pro	CTG-Leu-CTAL-Val-TCC-Ser-70	TAC Tyr CTC Leu GCC Ala 55 CAG Gln	TAT TYT  TTC Phe 40 TAC TYT  AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC Ile	ATG Met GGT Gly ACC Thr CAA Gln	GTT Val  CAG GIN  CTT Leu 75	a Gl GGG Gly GAC Asp AGG Arg 60 GTA Val	AAC AS GGG G1y 45 TTT Phe AAC AS GAC	TGG Trp 30 AGG Arg ACA Thr ACT Thr	GCT Ala ACC Thr TCC Ser AAC Asn	TP GI AAG Lys CAC His TTC Phe GGC Gly 80 AAC	97 145 193 241

ACC G	GG	TTC	CTT	GCC	GCG	CTG	TTC	TAC	ACC	CAC	AGC	TTC	AAC	GCG	TCC	GG	339
Thr G	Sly	Phe	Leu	Ala	Ala	Leu	Phe	Tyr	Thr	His	Ser	Ph	Asn	Ala	Ser		
			100					105					110				

SEQ ID NO:14

SEQUENCE LENGTH: 339 base pairs
SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
ANTI-SENSE: NO
ORIGINAL SOURCE

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE
CLONE: N27-2

C CGG ATC CCA CAA GCC GTG GTG GAT ATG GTG GCG GGG GCC CAC TGG GGA 49 Arg Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys GTC TTG GTT GTG ATG CTG CTT TTC GCC GGT GTT GAC GGG GGG ACC CAC Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His GTG ACA GGG GGG AAG GTA GCC TAC ACC ACC CAG AGC TTC ACA TCC TTC Val Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Ser Phe Thr Ser Phe TTT TCA CGA GGG CCG TCT CAG AGG ATC CAA CTT GTA AAC ACT AAC GGC Phe Ser Arg Gly Pro Ser Gln Arg Ile Gln Leu Val Asn Thr Asn Gly AGC TGG CAC ATC AAT AGG ACT GCC CTG AAT TGC AAT GAC TCC CTT AAC Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn ACC GGG TTC CTT GCC GCG CTG TTC TAC ACC CAC AGC TTC AAC GCG TCC GG 339 Thr Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser 

	SEQ ID NO:15
	SEQUENCE LENGTH: 339 base pairs
	SEQUENCE TYPE: nucleic acid
5	STRANDEDNESS: double
	TOPOLOGY: linear
	ANTI-SENSE: No
	ORIGINAL SOURCE
10	ORGANISM: Hepatitis C virus
	IMMEDIATE EXPERIMENTAL SOURCE
	CLONE: N27-3
15	C CGG ATC CCA CAA GCC GTG GTG GAT ATG GTG GCA GGG GCC CAC TGG GGA 49
	Arg Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly
	5 10 15
	GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG 97
20	Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys
	20 25 30
	GTC TTG GTT GTG ATG CTG CTC TTC GCC GGT GTT GAC GGG GGG ACC CAC 145
	Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His
25	35 40 45
	GTG ACA GGG GGG AAG GTA GCC TAC ACC CAG GGC TTT ACA CCC TTC 193
	Val Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Gly Phe Thr Pro Phe
30	50 55 60
30	TTT TCA CGA GGG CCG TCT CAG AAA ATC CAA CTT GTA AAC ACT AAC GGC 241
	Phe Ser Arg Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly
	75 80
35	AGC TGG CAC ATC AAT AGG ACT GCC CTC AAT TGC AAT GAC TCC CTT AAC 289
00	Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn
	ACC GGG TTC CTT CCC CTC CTC TTC TTC TAG ACC CAG ACC ACC
	Thr Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser
40	100 105 110
	110
	SEO ID NO:16

SEQ ID NO:16

SEQUENCE LENGTH: 393 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

	ANTI-S	ense	: No														
5	ORIGIN	AL S	OURC	E													
J	ORGANI	SM:	Hepa	titi	s C ·	viru	В										
	IMMEDI.	ATE :	EXPE	RIME	NTAL	sou	RCE										
	CLONE:	N19	-1														
10	•																
10	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTT	AAC	ACC	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Cys	Asp	Asp	Ser	Leu	Asn	Thr	Gly	Phe	Leu	Ala	Ala	
	1				5					10					15		
15	CTG	TTC	TAC	ACG	CAC	AGG	TTC	AAC	GCG	TCC	GGA	<b>TGT</b>	CCG	GAG	CGT	ATG	96
	Leu	Phe	Tyr	Thr	His	Arg	Phe	Asn	Ala	Ser	Gly	Cys	Pro	Glu	Arg	Met	
				20					25					30			
			TGC														144
20	Ala	Gly	Сув	Arg	Pro	Ile	Asp	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	
			35					40					45				
			GTT														192
	Thr.		Val	Val	Pro	Asn		Ser	Asp	Gln	Arg		Tyr	Сув	Trp	His	
25	ma.c	50	00m	<b>223</b>	000		55					60					
			CCT														240
	1yr 65	AId	Pro	Arg	PIO	70	GIĀ	116	Val	Pro		Ser	Gln	Val	Сув	_	
		Carc	መልጠ	mcc	mm.c		CCA	100		C mm	75	ama				80	200
30			TAT Tyr														288
	110	vai	.171	Cys	85	THE	FIO	Ser	PIO	90	VAI	vai	GTĀ	Thr	95	Авр	
	CGT	TTC	GGC	GCC		ACG	ጥልሮ	244	TYCIC		220	አልጥ	CAC	»CC		CTC	336
			Gly														330
35			2	100			-1-		105	011			U14	110	wob	741	
	CTA	CTC	CTC	AAC	AAC	ACA	CGG	CCG	CCG	CAG	GGC	AAC	TGG		GGT	TGT	384
			Leu														
			115				_	120					125		4	-2 -	
40	ACC	TGG	ATG				•										393
	Thr	Trp	Met														
		130															
45	SEQ ID	NO:	17														
	SEQUEN	CE LI	ENGTI	H: 39	93 ba	se p	paire	5									

93

50

SEQUENCE TYPE: nucleic acid

	STRAND	edne	SS:	doub	le												
	TOPOLO	GY:	line	ar													
5	ANTI-S	ense	: No														
	ORIGIN	AL S	OURC	E					•					•			
	ORGANI	SM:	Hepa	titi	s C	viru	8										
	IMMEDI	ATE :	EXPE	RIME	NTAL	SOU	RCE										
10	CLONE:	N19	-2														
															·		
	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTT	AAC	ACC	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Leu	Asn	Thr	Gly	Phe	Leu	Ala	Ala	
15	1				5					10					15		
			TAC														96
	Leu	Phe	Tyr	Thr	His	Arg	Phe	Asn	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25		÷			30			
20			TGC														144
	Ala	Ser	Сув	Arg	Pro	Ile	Asp	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	
			35					40					45				
			GTT														192
25	Thr		Val	Val	Pro	Asn		Ser	Asp	Gln	Arg		Tyr	Сув	Trp	His	
	<b>~</b> 10	50					55					60					
			CCT														240
		ATA	Pro	Arg	Pro		Gly	Ile	Val	Pro		Ser	Gln	Val	Cys	_	
30	65	CITIC	m s m	maa		70					75					80	
			TAT														288
	PIO	vaı	Tyr	Сув	Pne 85	Thr	Pro	ser	Pro	-	Val	Val	Gly	Thr		Авр	
	CCT	רואוורי	ccc	ccc		B.C.C	ma m		maa	90					95		
35			GGC Gly														336
	9		GIJ	100	FIO	INT	TÄT	ABII	11p	GIY	ABN	ABN		110	Авр	VAI	
	СТА	CTC	CTC		AAC	ACA	CGG	CCG		CAA	ccc	AAC			ccm	mcm	204
			Leu														384
40			115				9	120	110	<b>3.11</b>	GLY	non.	125	FIIG	Gry	Cys	
	ACC	TGG															393
		Trp												•			333
		130															
45																	

. 94

50

	SEQ ID	NO:	18		•												
	SEQUEN	CE L	engt	H: 3	93 b	<b>ase</b> ;	pair	ន									
5	SEQUEN	CE T	YPE:	nuc	leic	aci	đ										
5	STRAND	EDNE	ss: ·	doub	le												
	TOPOLO	GY:	line	ar													
	ANTI-S	ense	: No														
	ORIĞIN	AL S	OURC	E													
10	ORGANI	SM: 1	Hepa	titi	в С ч	viru	8										
	IMMEDI	ATE I	EXPE	RIME	NTAL	sou	RCE										
	CLONE:	N19	-3							•							
15																	
15	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTT	AAC	ACC	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Leu	Asn	Thr	Gly	Phe	Leu	Ala	Ala	
	ļ				5					10					15		
20																ATG	96
20	Leu	Phe	Tyr	Thr	His	Arg	Phe	Asn	Ala	Ser	Gly	Cys	Pro	Glu	Arg	Met	
				20					25					30			
				CGC											•		144
25	Ala	Ser		Arg	Pro	Ile	Asp	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	
23			35					40					45				
				GTG													192
	Thr		VAI	Val	Pro	Asn		Ser	Asp	Gln	Arg		Tyr	Сув	Trp	His	
30	MA.C	50	aam	<b>663</b>	000	maa	55		~			60					
-				CGA													240
	65	VIG	PLO	Arg	PIO	70	GIY	116	val	PIO		ттр	GIN	vaı	Сув		
	_	CTC	መልመ	TGC	mmc		CCX	200	COM		75	ama.	000			80	
35				Cys													288
		741	-1.	CYS	85	1111	FLO	Ser	FIO	90	val	AGI	GTÄ	THE	95	Авр	
	CGT	TTC	GGC	GCC		ACG	тат	AAC	TCC		220	<b>ም</b> ልል	GAG	»CG		CTC	336
				Ala													220
40	5		2	100			-1-		105	021	21011	ABIL	GIU	110	мър	AGI	
	СТА	CTC	CTC	AAC	AAC	ACA	CGG	CCG	_	CAA	ccr	AAC	TCC		CCT	<b>መረ</b> መ	384
				Asn													301
			115				9	120			1		125	- 110	7-1	~J.5	
45	ACC	TGG															393
	_	Trp						_									
		~						-									

SEQ ID NO:19

5	SEQUEN	CE L	ENGT	н: 3	93 b	ase	pair	8				,					
	SEQUEN	CE T	YPE:	nuc	leic	aci	d							٠			
	STRAND	EDNE	ss:	doub	le												
	TOPOLO	GY:	line	ar													
10	ANTI-S	ense	: No														
	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa	titi	s C	viru	8										
	IMMEDI.	ATE	EXPE	RIME	NTAL	SOU	RCE										
15	CLONE:	H19	-2														
													•				
																GCG	48
	Glu	Ala	Val	Asn	Сув	Авр	Asp	Ser	Leu	Gln	Thr	Gly	Phe	Leu	Ala	Ala	
20	1				5					10					15		
									GCA								96
	Leu	Phe	Tyr		His	Arg	Phe	Asn	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25					30			
25									TTC								144
	Ala	Ser		Arg	Pro	Ile	Ser		Phe	Ala	Gln	Ġly	Trp	Gly	Pro	Ile	
			35					40					45				
									GAC								192
30	Inr		AST	Val	Pro	Asp		Ser	Asp	Gln	Arg	Pro	Tyr	Cys	Trp	His	
	mag	50					55					60					
									GTA								240
	65	ATG	PIO	Arg	Pro		GIY	ITe	Val	Pro		Ser	Gln	Val	Сув	_	
35		CINC	mam	mcc.	mmo	70	~~				75					80	
									CCT								288
	110	Val	ıyı	Сув	85	THE	PIO	ser	Pro		Val	vaı	GTA	Thr		Asp	
10	CGT	uh∕au.	ccr	CCC		»CC	mac.	N.C.C	TGG	90	C00		~~~		95		
₩									Trp								336
	9	001	011	100	FIO	1111	TAT	THE	105	GIÀ	ATE	ABN	GIU		Авр	Val	
	СТА	СТС	Сфф	•	ממ	»CG	CCT	ccc	CCA	CAC	ccc	220	maa	110	com	<b></b>	204
45									Pro								384
<b>4</b> 5	<b></b>		115	1	-14-16			120	110	GIII	GTÄ	voll		Fue	GTĀ	Cys	
								120					125				

	ACC	TGG	YEA &	3													393
	Thr	Tr	Met	Ė.													373
		130	)														
5																	
	SEQ II	NO:	20														
	SEQUEN	ICE I	engi	TH: 3	393 £	ase	pair	:8									
	SEQUEN	ICE I	YPE:	nuc	cleic	aci	.d										
10	STRAND	EDNE	ess:	doub	ole												
	TOPOLO	GY:	line	ar													
	ANTI-S	ENSE	: No	•													
	ORIGIN	AL S	OURC	E													
15	ORGANI	SM:	Hepa	titi	в С	viru	8										
	IMMEDI	ATE	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	H19	-4				-	,									
											•						
20	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTC	CAG	ACT	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Leu	Gln	Thr	Gly	Phe	Leu	Ala	Ala	
	,1				5					10					15		
	CTG	TTC	TAC	AGG	CAC	AGG	TTC	AAC	GCA	TCC	GGG	TGC	CCA	GAA	CGC	ATG	96
25	Leu	Phe	Tyr	Arg	His	Arg	Phe	Asn	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25					30			
	GCC	AGC	TGT	CGC	CCC	ATT	AGC	GAG	TTC	GCT	CAG	GGG	TGG	GGC	CCT	ATC	144
	Ala	Ser	Сув	Arg	Pro	Ile	Ser	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	
30			35		_		_	40					45				
	ACT	CAT	GTT	GTG	CCT	GAC	GTG	TCG	GAC	CAG	AGG	CCT	TAT	TGC	TGG	CAC	192
	Thr	His	Val	Val	Pro	Asp	Val	Ser	Asp	Gln	Arg	Pro	Tyr	Сув	Trp	His	
		50					55					60					
35	TAC	GCG	CCT	CGA	CCG	TGC	GGT	ATC	GTA	CCC	GCG	TCG	CAG	GTG	TGT	GGT	240
	TYT	ALA	Pro	Arg	Pro		Gly	Ile	Val	Pro	Ala	Ser	Gln	Val	Сув	Gly	
	65	CITIC	<b></b>	<b>500</b>		70					75					80	
					TTC												288
40	Pro	vaı	туг	Сув	Phe	Thr	Pro	Ser	Pro		Val	Val	Gly	Thr	Thr	Asp	
	ccc	mcm	ccc	~~~	85					90					95		
	250	Cor	GGC	GCC	CCC	ACG	TAC	ACC	TGG	GGG	GCG	AAT	GAG	ACG	GAC	GTG	336
	ary	Ser	GIĀ		Pro	rnr	TYT	Thr		Gly	Ala	Asn	Glu		Asp	Val	
45	ርሞል	רישרי	Cthu	100	224	3.00	^~~	a==	105					110			
	CIM	CIC	CIT	MAC	AAC	ACG	CGT	CCG	CCA	CAG	GGC	AAC	TGG	TTC	GGT	TGT	384

	Leu	Leu	Leu	Asn	Asn	Thr	Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Сув	
			115					120					125				
_	ACC	TGG	ATG	;													393
5	Thr	Trp	Met	;													
		130	<b>,</b>														
10	SEQ ID																
10	SEQUEN						_	8									
	SEQUEN					aci	d										
	STRAND				le											•	
15	TOPOLO																
15	ANTI-S		-														
	ORIGIN			_													
	ORGANI																
20	IMMEDI			RIME	NTAL	SOU	RCE										
20	CLONE:	H19	-10														
	CNC	-	ama														
			GTG														48
25	1	Ala	Val	ABII	Cys 5	Asp	Авр	Ser	Leu		Thr	Gly	Phe	Leu		Ala	
		مامام	ጥልሮ	NGC.	_	N.C.C	mma			10					15		
			TAC Tyr														96
			-1-	20	1110	мц	File	ABII	25	ser	GTĀ	Cys	PTO		Arg	Met	
30	GCC	AGC	TGC		ccc	Ymm	) AGC	GAG	_	CCM	CNC		mca	30			
			Сув														144
			35				001	40	1116	ALG	GIII	GLY	45	GIY	PLO	11e	
	ACT	CAT	GTT	GTG	ССТ	GAC	GTG		GAC	CAG	AGG	CCT		TICC	TVCC	CAC	192
35			Val														13%
		50				•	55				9	60	-1-	CJB	*-P	HIO	
	TAC	GCA	CCT	CGA	CCG	TGC	GGT	GTC	GTA	ccc	GCG		CAG	GTG	TGT	CGT	240
			Pro														
40	65					70	-				75				-,-	80	
	CCA	GTG	TAT	TGC	TTC	ACC	CCA	AGC	CCT	GTC	GTG	GTG	GGG	ACG	ACC		288
			Tyr														
					85					90			-		95		
45	CGC	TCT	GGC	GCC	CCC	ACG	TAC	ACC	TGG	GGG	GCG	AAT	GAG	ACG	GAC	GTG	336
			Gly														
									_	_					-		

				100					10					11	0		
	CTA	CTC	CTI	C AAC	AA	CAC	G CG	T CC	G CC	A CA	G GG	C AA	C TG	G TT	C GG	T TGT	384
5	Leu	Let	ı Let	ı Asn	Ası	n Th:	r Ar	g Pro	Pr	o Gl	n Gl	y As:	n Tr	p Ph	e Gl	у Сув	
			115	•				120					12				
			ATC														393
	Thr	Trp	Met	:										•			
10	•	130	)														
70																	
	SEQ ID	NO:	22														
	SEQUEN	CE L	ENGT	н: 3	93 E	ase	pair	:8									
	SEQUEN	CE T	YPE:	nuc	leic	aci	ld									•	
15	STRAND	edne	SS:	doub	le												
	TOPOLO	GY:	line	ar													
	ANTI-S	ense	: No									*					
	ORIGIN	AL S	OURC	E													
20	ORGANIS	SM:	Hepa	titi	вС	viru	8										
	IMMEDIA																
	CLONE:																
25	GAG	GCC	GŤG	AAC	TGC	GAT	GAC	TCC	CTC	CAG	<b>ል</b> ሮሞ	ccc	mma	Com		GCG	40
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Ten.	Gln	mp x	Glu	Dho	CII Tom	31-	Ala	48
	1				- 5					10		GLY	FME	ren			
	CTG	TTC	TAC	AGG	CAC	AGG	TTC	AAC	CCA			<b>TOC</b> C	CCB	~~	15	ATG	
30	Leu	Phe	Tyr	Arg	His	Ara	Phe	Agn	Ala	202	C1	750	D	GAA	CGC	ATG	96
			=	20		5			25	261	GIY	Сув	Pro		Arg	Met	
	GCC	AGC	TGT	CGC	CCC	Arm	AGC	GAG		CCB	CNC	000	maa	30			
	Ala	Ser	Cvs	Arg	Pro	Tle	Ser	Glu	Dho	31-	CAG	GGG	166	GGC	CCT	ATC	144
35			35				Der	40	FIIE	wra	GIN	GIĀ		GIY	Pro	Ile	
	ACT	CAT		GTG	<del>ርር</del> ሞ	GAC	CTC		CNO	030			45				
	Thr	His	Val	Val	Pro	lan	Ual	Com	GAC	CAG	AGG	CCT	TAT	TGC	TGG	CAC	192
		50		Val		mp	55	SeT	мвр	GIN	Arg		Tyr	Сув	Trp	His	
40	TAC		<b>ሮ</b> ሮሞ	CGA	CCG	mcc		3 m.c	~~·			60					
	TAC	Ala	Pro	Ara	D~~	TGC.	GGT	ATC	GTA	ccc	GCG	TCG	CAG	GTG	TGT	GGT	240
	<b>Tyr</b> 65		-10	ar.y	LTO		GTĀ	TTE	val	Pro		Ser	Gln	Val	Cys	Gly	
		CTV	mam	mco ·	mun~	70	00-				75					80	
45	CCA (	Ua I	Meren.		LIC	ACC	CCA	AGC	CCT	GTC	GTG	GTG	GGG	ACG	ACC	GAT	288
<del>4</del> 0	Pro '	AGT	TAL	cys :		Thr	Pro	Ser	Pro		Val	Val	Gly	Thr	Thr	Asp	
					85					90					95		

	CGC	TCT	GGC	GCC	CCC	ACG	TAC	ACC	TGG	GGG	GCG	AAT	GAG	ACG	GAC	GTG	336
			Gly														
				100					105					110	•		
5	CTA	CTC	CTT	AAC	AAC	ACG	CGT	CCG	CCA	CAG	GGC	AAC	TGG	TTC	GGT	TGT	384
			Leu														
			115					120			_		125		•		
	ACC	TGG	ATG														393
10	Thr	Trp	Met														
		130															
	SEQ ID	NO:	23													•	
15	SEQUEN	CE L	engti	H: 3	93 ba	<b>ase</b> ]	pair	В									
	SEQUEN	CB T	YPE:	nuc	leic	aci	i										
	STRAND	EDNE	SS: c	doub	le												
	TOPOLO	GY:	linea	ar					•		,						
20	ANTI-S	ense	: No														
	ORIGIN	AL S	OURCI	3													
	ORGANI	5M: 1	Hepat	titi	8 C 1	virus	3										
	IMMEDIA	ATE I	EXPEI	RIME	NTAL	SOU	RCE										
25	CLONE:	¥19-	-6	_													
														•			
	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTC	CAG	ACT	GGG	TTC	CTT	GCC	ACG	48
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Leu	Gln	Thr	Gly	Phe	Leu	Ala	Thr	
30	. 1	-			5					10					15		
	CTG	TTC	TAC	AGG	CAC	AGG	TTC	AAC	GCA	TCC	GGG	TGC	CCA	GAA	CGC	ATG	96
	Leu	Phe	Tyr	Arg	His	Arg	Phe	Asn	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25					30			
35			TGT														144
	Ala	Ser	Cys	Arg	Pro	Ile	Ser	Glu	Phe	Ala	Gln	Gly	Trp	Asp	Pro	Ile	
			35					40					45				
			GTT														192
40	Thr	His	Val	Val	Pro	Asp	Val	Ser	Asp	Gln	Arg	Pro	Tyr	Сув	Trp	His	
		50					55					60			•		
			CCT														240
		Ala	Pro	Arg	Pro	Сув	Gly	Ile	Val	Pro	Ala	Ser	Gln	Val	Сув	Gly	
45	65					70					75					80	
	CCA	GTG	TAT	TGC	TTC	ACC	CCA	AGC	CCT	GTC	GTG	GTG	GGG	ACG	ACC	GAT	288
																•	

5

	Pro	Val	Tyr	Сув	Phe	Thr	Pro	Ser	Pro	Val	Val	Val	Gly	Thr	Thr	Asp	
					85					90					95		
5																GTG	336
	Arg	ser	GIĀ		Pro	Thr	Tyr	Thr		Gly	Ala	Asn	Glu	Thr	Авр	Val	
	0.00	0770		100					105					110			
						ACG											384
10	Leu	Leu		ASN	Asn	Thr	Arg		Pro	Gln	Gly	Asn		Phe	Gly	Cys	
	» CC	TGG	115					120					125				
			_														393
	THE	Trp	Met														
15		130															
	SEQ ID	NO:	24														
	SEQUEN			4. 30	13 h	100 1	na i re										
	SEQUEN					-		•									
20	STRAND					4020	•										
	TOPOLO																
	ANTI-S	ENSE	: No														
	ORIGIN	AL S	OURCE	3													
25	ORGANIS	5M: I	Tepat	itis	. C 1	irus	3										
	IMMEDIA	ATE I	ZXPEF	RIMEN	LAT	SOUF	CE							•			
	CLONE:	¥19-	-7										•				
30	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTC	CAG	ACT	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Сув	Авр	Asp	Ser	Leu	Gln	Thr	Gly	Phe	Leu	Ala	Ala	
	1				5					10					15		
						AGG											96
35	Leu	Phe	Tyr	Arg	His	Arg	Phe	qaA	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25					30			
						TTA											144
	Ala	Ser		Arg	Pro	Ile	Ser	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	
40			35		:			40					45				
						GAC											192
	Thr		Val	Val	Pro	Asp		Ser	Asp	Gln	Arg		Tyr	Сув	Trp	His	
	ma.c	50	~~~	003	000	<b></b>	55					60					
45						TGC											240
	TYE	WIG	PTO	w.d	PTO	Сув	дТĀ	TTE	Val	Pro	Ala	Ser	Gln	Val	Сув	Gly	

5

	65					70					75					80	
	CCA	GTG	TAT	TGC	TTC	ACC	CCA	AGC	CCT	GTC			GGG	ACG	ACC		288
5			Tyr														200
5					85					90			•		95		
	CGC	TCT	GGC	GCC	CCC	ACG	TAC	ACC	TGG	GGG	GCG	AAT	GAG	ACG	GAC	GTG	336
	Arg	Ser	Gly	Ala	Pro	Thr	Tyr	Thr	Trp	Gly	Ala	Asn	Glu	Thr	Авр	Val	
10	•			100					105					110			
			CTT														384
	Leu	Leu	Leu	Asn	Asn	Thr	Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Cys	
			115					120					125				
15		TGG	_													•	393
	rnr	Trp	Met														
		130															
	SEQ ID	NO - 1	25														
20	SEQUEN		_	1: 6:	99 hr		na i r										
	SEQUEN					_		5					•				
	STRAND																
	TOPOLO	GY: ]	linea	ır													
25	ANTI-SI	ENSE:	No														
	ORIGIN	AL SO	OURCE	3													
	ORGANIS	SM: F	lepat	itis	C	rirus	3										
	IMMEDIA	ATE I	ZXPER	IMEN	ITAL	SOUP	RCE										
30	CLONE:	MX24	1-4														
			CGG														48
35		Thr	Arg	Pro		Gln	Gly	Asn	Trp	Phe	Gly	Сув	Thr	Trp	Met	Asn	
33	1		~~~		5					10					15		
			GGG														96
	GIŢ	THE	Gly	Pne 20	Thr	гÀв	Thr	Сув		Gly	Pro	Pro	Сув		Ile	Gly	
40	GGG	GTC	GGC	-	חממ	NCC.	mmc	» Cm	25 mag			<b>a.</b> a		30			
			Gly														144
	1		35	21011	71011	1111	Deu	40	СУВ	PIO	THE	Авр		Pne .	Arg	rås	
	CAC	ccc	GAG	GCC	ACT	TAC	ACA		ጥርጥ ተ	ርርጥ	ጥርር	GGG	45 CCT	ጥርረ	ጥብረ	<b>ACC</b>	102
45			Glu														192
		50				-	55	•	- 4 -	1		60		F			

	CCT	AGG	TGC	CTA	GTT	CAT	TAC	CCA	TAC	AGG	CTC	TGG	CAC	TAT	CCC	TGC	240
	Pro	Arg	Сув	Leu	Val	His	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	
5	65					70					75					80	
	ACT	GTC	AAC	TTT	ACC	ATC	TTC	AAG	GTT	AGG	ATG	TAT	GTG	GGG	GGC	GTG	288
	Thr	Val	Asn	Phe	Thr	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	
					85					90					95		
10	GAA	CAC	AGG	CTT	GAA	GCT	GCA	TGC	AAT	TGG	ACC	CGA	GGA	GAG	CGT	TGT	336
70	Glu	His	Arg	Leu	Glu	Ala	Ala	Сув	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	
				100					105					110			
	GAC	TTG	GAG	GAC	AGG	GAT	AGA	TCA	GAG	CTT	AGC	CCG	CTA	TTG	CTG	TCC	384
15	Asp	Leu	Glu	Asp	Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	
75			115					120					125				
	ACA	ACA	GAG	TGG	CAG	GTA	CTG	CCC	TGT	TCC	TTC	ACC	ACC	CTG	CCG	GCT	432
	Thr	Thr	Glu	Trp	Gln	Val	Leu	Pro	Сув	Ser	Phe	Thr	Thr	Leu	Pro	Ala	
00		130					135					140					
20	CTG	TCC	ACT	ggt	TTG	ATT	CAT	CTC	CAT	CAG	AAC	ATC	GTG	GAC	GTG	CAA	480
	Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	
	145					150					155					160	
05	TAT	CTG	TAC	GGC	ATA	GGG	TCG	GCG	GTT	GTC	TCC	TTC	GCA	ATC	AAA	TGG	528
25	Tyr	Leu	Tyr	Gly	Ile	Gly	Ser	Ala	Val	Val	Ser	Phe	Ala	Ile	Lys	Trp	
					165					170					175		
						CTT											576
20	Glu	Tyr	Ile	Leu	Leu	Leu	Phe	Leu	Leu	Leu	Ala	ysb	Ala	Arg	Val	Сув	
30				180					185					190			
	GCC	TGC	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAC	GCC	GAC	GCC	ACC	TTA	624
	Ala	Cys	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	His	Ala	Asp	Ala	Thr	Leu	
			195					200					205				
35	GAG	AA															629
	Glu																

SEQ ID NO:26

SEQUENCE LENGTH: 629 base pairs
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

45 ANTI-SENSE: No

ORIGINAL SOURCE

55

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE

CLONE: MX24-5

																	•
	AAC	ACA	CGG	CCG	CCG	CAG	GGG	AAC	TGG	TTT	GGC	TGT	ACA	TGG	ATG	AAT	48
	Asn	Thr	Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Сув	Thr	Trp	Met	Asn	
10	1				5					10					15		
70	GGC	ACT	GGG	TTC	ACA	AAG	ACG	TGC	GGG	GGC	CCC	CCG	TGC	AAC	ATC	GGG	96
	Gly	Thr	Gly	Phe	Thr	Lys	Thr	Сув	Gly	Gly	Pro	Pro	Сув	Asn	Ile	Gly	
				20					25					30			
45	GGG	GTC	GGC	AAC	AAT	ACC	TTG	ACT	TGC	CCC	ACG	GAC	TGC	TTC	CGG	AAG	144
15	Gly	Val	Gly	Asn	Asn	Thr	Leu	Thr	Сув	Pro	Thr	Авр	Сув	Phe	Arg	Lys	
			35					40					45		,		
	CAC	CCC	GAG	GCC	ACT	TAC	ACA	AAA	TGT	<b>GGT</b>	TCG	GGG	CCT	TGG	TTG	ACG	192
	His	Pro	Glu	Ala	Thr	Tyr	Thr	Lys	Сув	Gly	Ser	Gly	Pro	Trp	Leu	Thr	•
20		50					55					60					
	CCT	AGG	TGC	CTA	GTT	CAT	TAC	CCA	TAC	AGG	CTC	TGG	CAC	TAT	CCC	TGC	240
	Pro	Arg	Сув	Leu	Val	His	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Сув	
	65					70					75					80	
25	ACT	GTC	AAC	TTT	ACC	ATC	TTC	AAG	GTT	AGG	ATG	TAT	GTG	GGG	GGC	GTG	288
	Thr	Val	Asn	Phe	Thr	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	
					85					90			-		95		
	GAA	CAC	AGG	CTT	GAA	GCT	GCA	TGC	AAT	TGG	ACC	CGA	GGA	GAG	CGT	TGT	336
30	Glu	His	Arg	Leu	Glu	Ala	Ala	Сув	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	
				100					105					110			
	GAC	TTG	GAG	GAC	AGG	GAT	AGA	TCA	GAG	CTT	AGC	CCG	CTA	TTG	CTG	TCT	384
	Asp	Leu	Glu	Asp	Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	
35			115					120					125				
															CCG		432
	Thr	Thr	Glu	Trp	Gln	Val	Leu	Pro	Cys	Ser	Phe	Thr	Thr	Leu	Pro	Ala	
		130					135					140					
40	CTG	TCC	ACT	GGT	TTG	ATT	CAT	CTC	CAT	CAG	AAC	ATC	GTG	GAC	GTG	CAA	480
	Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	qaA	Val	Gln	
	145					150					155					160	
	TAT	TTG	TAC	GGC	ATA	GGG	TCG	GCG	GTT	GTC	TCC	TTC	GCA	ATC	AAA	TGG	528
45	Tyr	Leu	Tyr	Gly	Ile	Gly	Ser	Ala	Val	<b>Val</b>	Ser	Phe	Ala	Ile	Lys	Trp	
					165					170					175		
				-													

50

	GAA	TAT	ATT	CTG	TTG	CTT	TTC	CTT	CTC	CTG	GCG	GAC	GCG	CGC	GTC	TGT	576
								Leu									
_				180					185					190			
5	GCC	TGC	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAC	GCC	GAC	GCC	ACC	TTA	624
	Ala	Сув	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	His	Ala	Авр	Ala	Thr	Leu	-
		•	195					200					205				
	ĠAG	AA	ė											•			629
10	Glu																
	SEQ ID	NO:	27														
	SEQUEN	CE L	ENGT	H: 6:	29 ba	ase 1	pair	8									
15	SEQUEN						•									•	
	STRAND	EDNE	SS: d	doub	le												
	TOPOLO	GY: 3	line	ar													
	ANTI-S	ense	: No														
20	ORIGIN	AL S	OURC	E													
	ORGANI	SM: I	Hepat	titia	5 C 1	viru	3										
	IMMEDIA	ATE I	EXPEI	RIMEI	NTAL	SOUI	RCE										
	CLONE:	MX2	4-13														
25																	
	AAC	ACA	CGG	CCG	CCG	CAG	GGG	AAC	TGG	TTT	GGC	TGT	ACA	TGG	ATG	AAT	48
	Asn	Thr	Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Сув	Thr	Trp	Met	Asn	
	1				5					10					15		
30								TGC									96
	Gly	Thr	Gly	Phe	Thr	Lys	Thr	Сув	Gly	Gly	Pro	Pro	Сув	Asn	Ile	Gly	
				20					25					30			
								ACT									144
35	Gly	Val		naA	Asn	Thr	Leu	Thr	Сув	Pro	Thr	Asp	Сув	Phe	Arg	Lys	
			35					40					45				
								AAA									192
	His		Glu	Ala	Thr	Tyr	Thr	Lys	Сув	Gly	Ser	Gly	Pro	Trp	Leu	Thx	
40	_4_	50					55	•				60					
								CCA									240
		Arg	Cys	Leu	Val		Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Сув	
	65	052	• • • •			70	<u> </u>				75					80	
45								AAG									288
	Thr	Val	Asn	Phe	Thr	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	

					85					90					95		
	GAA	CAC	AGG	CTT	GAA	GCT	GCA	TGC	AAT	TGG	ACC	CGC	GGA	GAG	CGT	TGT	336
	Glu	His	Arg	Leu	Glu	Ala	Ala	Сув	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Сув	
5				100					105					110			
	GAC	TTG	GAG	GAC	AGG	GAT	AGA	TCA	GAG	CTT	AGC	CCG	CTA	TTG	CTG	TCT	384
	Asp	Leu	Glu	Asp	Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	
	•		115					120					125				
10	ACA	ACA	GAG	TGG	CAG	GTA	CTG	CCC	TGT	TCC	TTC	ACC	ACC	CTG	CCG	GCT	432
	Thr	Thr	Glu	Trp	Gln	Val	Leu	Pro	Сув	Ser	Phe	Thr	Thr	Leu	Pro	Ala	
		130					135					140					
	CTG	TCC	ACT	GGT	TTG	ATT	CAT	CTC	CAT	CAG	AAC	ATC	GTG	GAC	GTG	CAA	480
15	Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	
	145					150					155					160	
	TAT	CTG	TAC	GGC	ATA	GGG	TCG	GCG	GTT	GTC	TCC	TTC	GCA	ATC	AAA	TGG	528
	Tyr	Leu	Tyr	Gly	Ile	Gly	Ser	Ala	Val	Val-	Ser	Phe	Ala	Ile	Lys	Trp	
20					165					170					175		
	GAA	TAT	ATT	CTG	TTG	CTT	TTC	CTT	CTC	CTG	GCG	GAC	GCA	CGC	GTC	TGT	576
	Glu	Tyr	Ile	Leu	Leu	Leu	Phe	Leu	Leu	Leu	Ala	Asp	Ala	Arg	Val	Сув	
				180					185					190			
25	GCC	TGC	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAC	GCC	GAC	GCC	ACC	TTA	624
	Ala	Суз	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	His	Ala	Авр	Ala	Thr	Leu	
			195					200					205	•			
	GAG	AA															629
30	Glu																

SEQ ID NO:28

SEQUENCE LENGTH: 652 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: No

40 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

CLONE: N27N19-1

C CGG ATC CCA CAA GCC GTG GTG GAT ATG GTG GCA GGG GCC CAC TGG GGA 49

55

	A	rg I	le P	ro G	ln A	la V	al V	al A	sp M	et V	al A	la G	ly A	la H	is T	rp Gly	•
		1				5					10					15	
5	GTC	CTG	GCG	GGC	CTT	GCC	TAC	TAT	TCC	ATG	GTG	GGG	AAC	TGG	GCT	AAG	97
•	Val	Leu	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	
				20					25					30			
	GTC	TTG	GTT	GTG	ATG	CTG	CTC	TTC	GCC	GGT	GTT	GAC	GGG	GGG	ACC	CAC	145
10	Val	Leu	Val	Val	Met	Leu	Leu	P(e	Ala	GLy	VAl	Asp	Gly	Gly	THr	His	
			35					40					45				
	GTG	ACA	GGG	GGG	AAG	GTA	GCC	TAC	ACC	ACC	CAG	GGC	TTT	ACA	CCC	TTC	193
	Val	Thr	Gly	Gly	Lys	Val	Ala	Tyr	Thr	Thr	Gln	Gly	Phe	Thr	Pro	Phe	
15		50					55					60	:				
15									ATC								241
	Phe	Ser	Arg	Gly	Pro	Ser	Gln	Lys	Ile	Gln	Leu	Val	Asn	Thr	Asn	Gly	
	65					70					75					80	
20	AGC	TGG	CAC	ATC	AAT	AGG	ACT	GCC	CTC	AAT	TGC	AAT	GAC	TCC	CTT	AAC	289
20	Ser	Trp	His	Ile	Asn	Arg	Thr	Ala	Leu	Asn	Сув	Asn	Asp	Ser	Leu	Asn	
					85					90					95		
									TAC								337
	Thr	Gly	Phe	Leu	Ala	Ala	Leu	Phe	Tyr	Thr	His	Ser	Phe	Asn	Ala	Ser	
25				100					105					110			
									TGC								385
	Gly	Суз	Pro	Glu	Arg	Met	Ala	Gly	Сув	Arg	Pro	Ile	Авр	Glu	Phe	Ala	
			115					120					125				
30									GTT								433
	Gln		Trp	Gly	Pro	Ile	Thr	His	Val	Val	Pro	Asn	Ile	Ser	Asp	Gln	
		130					135					140					
									CCT								481
35		Pro	Tyr	Сув	Trp	His	Tyr	Ala	Pro	Arg	Pro	Сув	Gly	Ile	Val	Pro	
	145					150					155					160	
									TAT								529
	Ala	Ser	Gln	Val	Сув	Gly	Pro	Val	Tyr	Сув	Phe	Thr	Pro	Ser	Pro	Val	
40					165					170					175		
									GGC								577
	Val	Val	Gly	Thr	Thr	Asp	Arg	Phe	Gly	Ala	Pro	Thr	Tyr	Asn	Trp	Gly	
				180					185					190			
45	AAC	TAA	GAG	ACG	GAT	GTG	CTA	CTC	CTC	AAC	AAC	ACA	CGG	CCG	CCG	CAG	625
	Asn	Asn	Glu	Thr	Asp	Val	Len	T.e.11	T.OII	Agn	λan	ሞትሎ	Ara	Dro	Dro	Gln.	

			195		-			200		•			205	i			
	GGC	AAC	TGG	TTC	GGT	TGT	ACC	TGG	ATG							•	652
5	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met								
3		210					215										
	SEQ ID	NO:	29														
10	SEQUEN	CE L	engt	H: 9	77 b	ase )	pair	6									
10	SEQUEN	CE T	YPE:	nuc	leic	aci	d										
	STRAND	EDNE	SS:	doub.	le												
	TOPOLO	GY:	line	ar									•				
	ANTI-S	ense	: No														
15	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa [.]	titi	вС	viru	5										
	IMMEDI.	ATE :	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	N19	MX 24	A-1							٠						
20										•							
	GAG	GCC	GTG	AAC	TGC	GAT	GAC	TCC	CTT	AAC	ACC	GGG	TTC	CTT	GCC	GCG	48
	Glu	Ala	Val	Asn	Сув	Asp	Asp	Ser	Leu	Asn	Thr	Gly	Phe	Leu	Ala	Ala	
	1				5					10					15		
25	CTG	TTC	TAC	ACG	CAC	AGG	TTC	AAC	GCG	TCC	GGA	TGT	CCG	GAG	CGT	ATG	96
	Leu	Phe	Tyr	Thr	His	Arg	Phe	Asn	Ala	Ser	Gly	Сув	Pro	Glu	Arg	Met	
				20					25					30			
	GCC	GGT	TGC	CGC	CCC	ATT	GAC	GAG	TTC	GCT	CAG	GGG	TGG	GGT	CCC	ATC	144
30	Ala	Gly	Сув	Arg	Pro	Ile	Asp	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	•
			35					40					45				
	ACT	CAT	GTT	GTG	CCT	AAC	ATC	TCG	GAC	CAG	AGG	CCC	TAT	TGC	TGG	CAC	192
	Thr	His	Val	Val	Pro	Asn	Ile	Ser	qaA	Gln	Arg	Pro	Tyr	Cys	Trp	His	
35		50					55					60					
	TAC	GCG	CCT	CGA	CCG	TGT	GGT	ATC	GTA	CCC	GCG	TCG	CAG	GTG	TGT	GGT	240
÷.		Ala	Pro	Arg	Pro	Сув	Gly	Ile	Val	Pro	Ala	Ser	Gln	Val	Сув	Gly	
	65					70					75					80	
40	CCG	GTG	TAT	TGC	TTC	ACC	CCA	AGC	CCT	GTT	GTG	GTG	GGG	ACG	ACC	GAT	288
	Pro	Val	Tyr	Cys	Phe	Thr	Pro	Ser	Pro	Val	Val	Val	Gly	Thr	Thr	Asp	
					85					90					95		
				GCC													336
45	Arg	Phe	Gly	Ala	Pro	Thr	Tyr	Asn	Trp	Gly	Asn	Asn	Glu	Thr	Авр	Val	
				100					105					110			

	CTA	CTC	CTC	AAC	AAC	ACA	CGG	CCG	CCG	CAG	GGC	AAC	TGG	TTC	GGT	TGT	384
	Leu	Leu	Leu	Asn	Asn	Thr	Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Сув	
5			115					120					125				
3	ACC	TGG	ATG	AAT	GGC	ACT	GGG	TTC	ACA	AAG	ACG	TGC	GGG	GGC	CCC	CCG	432
	Thr	Trp	Met	Asn	Gly	Thr	Gly	Phe	Thr	Lys	Thr	Сув	Gly	Gly	Pro	Pro	
		130					135					140					
10	TGC	AAC	ATC	GGG	GGG	GTC	GGC	AAC	AAT	ACC	TTG	ACT	TGC	CCC	ACG	GAC	480
10	Сув	Asn	Ile	Gly	Gly	Val	Gly	Asn	Asn	Thr	Leu	Thr	Сув	Pro	Thr	Asp	
	145					150					155					160	
	TGC	TTC	CGG	AAG	CAC	CCC	GAG	GCC	ACT	TAC	ACA	AAA	TGT	<b>GGT</b>	TCG	GGG	528
15	Сув	Phe	Arg	Lys	His	Pro	Glu	Ala	Thr	Tyr	Thr	Lys	Сув	Gly	Ser	Gly	
					165					170		•			175		
	CCT	TGG	TTG	ACG	CCT	AGG	TGC	CTA	GTT	CAT	TAC	CCA	TAC	AGG	CTC	TGG	576
	Pro	Trp	Leu	Thr	Pro	Arg	Cys	Leu	Val	His	Tyr	Pro	Tyr	Arg	Leu	Trp	
20				180					185		-			190			
20	CAC	TAT	CCC	TGC	ACT	GTC	AAC	TTT	ACC	ATC	TTC	AAG	GTT	AGG	ATG	TAT	624
	His	Tyr	Pro	Сув	Thr	Val	Asn	Phe	Thr	Ile	Phe	Lys	Val	Arg	Met	Tyr	
			195					200					205				
25	GTG	GGG	GGC	GTG	GAA	CAC	AGG	CTT	GAA	GCT	GCA	TGC	AAT	TGG	ACC	CGA	672
20	Val	_	Gly	Val	Glu	His	Arg	Leu	Glu	Ala	Ala	-	Asn	Trp	Thr	Arg	
		210					215					220					
			-			_	GAG					-		-			720
30		Glu	Arg	Сув	Asp		Glu	Asp	Arg	qaA		Ser	Glu	Leu	Ser		
00	225					230					235					240	
							GAG										768
	Leu	Leu	Leu	Ser		Thr	Glu	Trp	Gln		Leu	Pro	Сув	Ser		Thr	
35		ama	~~~		245	<b></b>		00m		250	a	ama	~~ m	a.a	255	<b>&gt; 20</b>	016
	_			_			ACT	_		_				_			816
	The	ren	PIO		Leu	Ser	Thr	GIY		TTG	H18	rea	H18		ABN	116	
	CMC	CNO	ama	260	mam	cmc	ma.c	000	265	000	maa	000	com	270	maa	mmc.	064
40							TAC										864
	Val	Авр		GIII	TAL	reu	Tyr		110	GTĀ	Ser	WIG		vai	Set	PHE	
	CCA	3.000	275	mcc	C3.3	mam.	3 (1101)	280 cmc	mmc	com.	mma	ama	285	omo	000	CAC	912
							ATT										312
45	WIG		тув	rrb	GIU	TAL	11e 295	neu	rea	Tea	FIIG		TRU	ned	wra	wah	
70	CCC	290	CMC	un, om	ccc	mcc		ጥርር	A mv	אוווער	Curc	300	am a	GCC	CAC	CCC	960
	GCG	CGC	GTC	IGT	GCC	160	TTG	100	ATG	ATG	CIG	CTG	ATA	GUU	UAC	GCC	300

	Ala	Arg	Val	Cys	Ala	Сув	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	His	Ala	
	305					310					315					320	
5	GAC	GCC	ACC	TTA	GAG	AA											977
·	Asp	Ala	Thr	Leu	Glu												
					325									•		•	
10	SEQ ID	NO:	30														
10	SEQUEN	CE L	ENGT	H: 9	77 b	ase 1	pair	8									
	SEQUEN																
	STRAND												•				
	TOPOLO																
15	ANTI-S																
	ORIGIN																
	ORGANI				вС	vi <del>r</del> nı	Ri										
	IMMEDI																
20	CLONE:					000.											
																	,
	GAG	GCC	GTG	AAC	TGC	САТ	GAC	ሞሮር	Сфф	AAC	»CC	ccc	тите	Custo	ccc	ccc	40
		Ala															48
25	1				5		-mp	561	Tou	10	1111	GTĀ	FIIG	reu	15	Ala	
		TTC	TAC	ACG	_	ACC	كىلمل	AAC	GCG		CCA	anca an	ccc	CAC		3.000	0.0
		Phe															96
			-1-	20		-u-y	1116	Aon	25	Set	GIY	Cys	PIO	30	Arg	Met	
30	GCC	GGT	ጥርር		cec	Valua	GAC	GNG	_	CCM	CNC	ccc	mcc.		000	****	144
		Gly															144
		027	35	9	110	110	vob	40	rne	VIG	GIII	GTĀ	45	GTĀ	Pro	116	
	АСТ	CAT		GTG	ርርሞ	AAC	ልጥሮ		CAC	CAC	NCC.	aaa.		maa.	maa	<b>a</b> 20	100
35		His															192
		50	741	V4.1	110	non	55	261	мвр	GTII	Arg		ıyr	Сув	тгр	HIS	
	ሞልሮ	GCG	ር ር	CGA	ccc	መረያጥ		አመሮ	CMA	000	000	60 mag	~~	ama	mam		040
		Ala															240
40	65		110	y	110	70	GIY	116	Val	PLO		ser	GIN	Val	Сув	_	
		GTG	ጥልጥ	שכיר	THE THE		CCA	300	oom.		75	~~~				80	
																	288
	110	Val	TÄT	Cyb	85	THE	PIO	Ser	PIO		Val	vaı	GTĀ	Thr		Asp	
45	CCT	mmc	ccc	CCC		100	m> c			90					95		
		TTC															336
	ALY.	Phe	GTĀ	WIG	PIO	THE	TÄĽ	Asn	Trp	GIY	Asn	Asn	Glu	Thr	Asp	Val	
50																	
<b>J</b> U																	

				100													
	Cmy	CMC	CMC				222		105					110			
			CTC														384
5	nea	red	Leu	Asn	Asn	Tnr	Arg		Pro	Gln	Gly	Asn		Phe	Gly	Cys	
	a C a	mcc.	115	2 2 0	000	» om	000	120					125				
			ATG														432
		130	Met	Asn	GIY	Thr		hue	Thr	Lys	Thr		Gly	Gly	Pro	Pro	
10	ሞርር		አመሮ	ccc	ccc	CMC	135					140					
			ATC														480
	145	non	Ile	GIY	Gry	150	GIY	ASI	ASI	Thr		Inr	Cys	Pro	Thr		
		րու	CGG	AAG	CAC		CAC	ccc	3 Cm	ma o	155		mam	aam		160	
15			Arg														528
	-1-		9	_, _	165	110	GIU	VIG	THE	170	THE	цув.	Cys	GIÀ		GIÀ	
	CCT	TGG	TTG	ACG		AGG	ጥርር	ርጥል	Cour		ሞልሮ	CCA	መአሮ	NCC.	175	mcc	57 <i>6</i>
20			Leu														576
		_		180		5	-1-		185			110	-1-	190	nea	rrp	
	CAC	TAT	CCC	TGC	ACT	GTC	AAC	TTT		ATC	TTC	AAG	GTT		ATG	ТАТ	624
			Pro														
25			195					200				-	205	_		•	
	GTG	GGG	GGC	GTG	GAA	CAC	AGG	CTT	GAA	GCT	GCA	TGC	AAT	TGG	ACC	CGA	672
	Val	${\tt Gly}$	Gly	Val	Glu	His	Arg	Leu	Glu	Ala	Ala	Cys	Asn	Trp	Thr	Arg	
		210					215					220					
30	GGA	GAG	CGT	TGT	GAC	TTG	GAG	GAC	AGG	GAT	AGA	TCA	GAG	CTT	AGC	CCG	720
	Gly	Glu	Arg	Сув	Asp	Leu	Glu	Asp	Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	
	225					230					235					240	
35			CTG														768
00	Leu	Leu	Leu	Ser		Thr	Glu	Trp	Gln	Val	Leu	Pro	Cys	Ser	Phe	Thr	
					245					250					255		
			CCG														816
40	Thr	Leu	Pro		Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	
	Oma.	0.0	<b>a</b> ma	260					265					270			
			GTG														864
	vai	Asp	Val	GIN	Tyr	Leu	Tyr		Ile	Gly	Ser	Ala		Val	Ser	Phe	
45	CCX	» mc	275	maa	<b>63.3</b>			280					285				
			AAA														912
	vra		Lys	Trp	GIU	ıyr		ьеп	Leu	Leu	Phe		Leu	Leu	Ala	Asp	
		290					295					300					

	GCG	CGC	GTC	TGT	GCC	TGC	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAC	GCC	960
	Ala	Arg	Val	Cys	Ala	Cys	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	His	Ala	
_	305					310					315					320	
5	GAC	GCC	ACC	TTA	GAG	AA											977
	Asp	Ala	Thr	Leu	Glu												
	•				325												
10																	
	SEQ ID	NO:	31														
	SEQUENC	E L	engti	H: 1	236 1	oase	pai:	rs									
	SEQUENC	E T	YPE:	nuc	leic	aci	i					*					
15	STRANDE	DNE	SS: c	loub.	le												
	TOPOLOG	Y: :	linea	ar													
	ANTI-SE	ENSE	: No														
	ORIGINA	L S	OURCE	3													
20	ORGANIS	M: 1	Hepat	titis	3 C 1	/irus	3						•				
	IMMEDIA	TE 1	EXPE	RIME	JATN	sou	RCE										
	CLONE:	N271	1X24 <i>1</i>	<b>1</b> -1													
25	C CC																
	C C6	G A.	rc co	CA C	AA GO	CC G	rg g	rg G	AT A	rg gr	rg go	CA GO	G G	CC CI	AC TO	G GGA	49
																G GGA	
	Ar	g II	le Pı	co Gl	ln Al	la Va 5	al Va	al As	sp Me	et Va	al Al LO	la Gi	ly Al	la Hi	is T	тр Gly 15	
	Ar GTC	g II 1 CTG	le Pi GCG	co GI	ln Al	la Va 5 GCC	al Va TAC	al As TAT	FP Me	et Va 1 ATG	al Al LO GTG	La GI GGG	Ly Al	la Hi TGG	is Ti	cp Gly 15 AAG	
30	Ar	g II 1 CTG	le Pi GCG	co GI	ln Al	la Va 5 GCC	al Va TAC	al As TAT	FP Me	et Va 1 ATG	al Al LO GTG	La GI GGG	Ly Al	la Hi TGG	is Ti	cp Gly 15 AAG	•
	Ar GTC Val	g II 1 CTG Leu	GCG Ala	GGC Gly 20	ln Al CTT Leu	la Va 5 GCC Ala	TAC	al As TAT Tyr	TCC Ser 25	et Va I ATG Met	al Al 10 GTG Val	da Gi GGG Gly	AAC Asn	TGG Trp 30	GCT	rp Gly 15 AAG Lys	•
	GTC Val GTC	rg II 1 CTG Leu TTG	GCG Ala	GGC Gly 20 GTG	CTT Leu ATG	la Va 5 GCC Ala CTG	TAC TYT	TAT TYT TTC	TCC Ser 25 GCC	ATG Met	al Al LO GTG Val	GGG GLY GAC	AAC Asn GGG	TGG Trp 30 GGG	GCT Ala	ep Gly 15 AAG Lys CAC	•
	Ar GTC Val	rg II 1 CTG Leu TTG	GCG Ala	GGC Gly 20 GTG	CTT Leu ATG	la Va 5 GCC Ala CTG	TAC TYT	TAT TYT TTC	TCC Ser 25 GCC	ATG Met	al Al LO GTG Val	GGG GLY GAC	AAC Asn GGG	TGG Trp 30 GGG	GCT Ala	ep Gly 15 AAG Lys CAC	97
	GTC Val GTC Val	rg II I CTG Leu TTG Leu	GCG Ala GTT Val	GGC Gly 20 GTG Val	CTT Leu ATG Met	5 GCC Ala CTG Leu	TAC Tyr CTC Leu	TAT TYT TTC Phe 40	TCC Ser 25 GCC Ala	ATG Met GGT	al Al iO GTG Val GTT Val	GGG GLY GAC Asp	AAC Asn GGG Gly 45	TGG Trp 30 GGG Gly	GCT Ala ACC	rp Gly 15 AAG Lys CAC His	97
30	GTC Val GTC Val	rg II  CTG  Leu  TTG  Leu  ACA	GCG Ala GTT Val 35 GGG	GGC Gly 20 GTG Val	CTT Leu ATG Met	5 GCC Ala CTG Leu	TAC Tyr CTC Leu GCC	TAT TYT TTC Phe 40 TAC	TCC Ser 25 GCC Ala	ATG Met GGT Gly	al Al lo GTG Val GTT Val	GGG GLY GAC Asp GGC	AAC Asn GGG Gly 45	TGG Trp 30 GGG Gly	GCT Ala ACC Thr	TTC	97
30	GTC Val GTC Val	TTG Leu ACA	GCG Ala GTT Val 35 GGG	GGC Gly 20 GTG Val	CTT Leu ATG Met	5 GCC Ala CTG Leu	TAC Tyr CTC Leu GCC	TAT TYT TTC Phe 40 TAC	TCC Ser 25 GCC Ala	ATG Met GGT Gly	al Al lo GTG Val GTT Val	GGG GLY GAC Asp GGC	AAC Asn GGG Gly 45	TGG Trp 30 GGG Gly	GCT Ala ACC Thr	TTC	97
30	GTC Val GTC Val GTG Val	TTG Leu ACA Thr 50	GCG Ala GTT Val 35 GGG Gly	GGC Gly 20 GTG Val GGG Gly	CTT Leu ATG Met AAG Lys	ta Va 5 GCC Ala CTG Leu GTA Val	TAC TYT CTC Leu GCC Ala 55	TAT TYT TTC Phe 40 TAC TYT	TCC Ser 25 GCC Ala ACC	ATG Met GGT Gly ACC	ol Al GTG Val GTT Val CAG Gln	GGG GLY GAC Asp GGC GLY 60	AAC Asn GGG Gly 45 TTT Phe	TGG Trp 30 GGG Gly ACA	GCT Ala ACC Thr	rp Gly 15 AAG Lys CAC His TTC	97
30	GTC Val GTC Val GTG Val	TTG Leu ACA Thr 50	GCG Ala GTT Val 35 GGG Gly	GGC Gly 20 GTG Val GGG Gly GGG	CTT Leu ATG Met AAG Lys	ta Va 5 GCC Ala CTG Leu GTA Val	TAC TYT CTC Leu GCC Ala 55 CAG	TAT TYT TTC Phe 40 TAC TYT	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr	ol Allio GTG Val GTT Val CAG Gln	GGG GLY GGC GLY GGC GLY GGTA	AAC Asn GGG Gly 45 TTT Phe	TGG Trp 30 GGG Gly ACA Thr	GCT Ala ACC Thr CCC Pro	TTC Phe GGC	97
30 35	GTC Val GTC Val GTG Val TTT Phe	TTG Leu ACA Thr 50	GCG Ala GTT Val 35 GGG Gly	GGC Gly 20 GTG Val GGG Gly GGG	CTT Leu ATG Met AAG Lys	ta Va 5 GCC Ala CTG Leu GTA Val	TAC TYT CTC Leu GCC Ala 55 CAG	TAT TYT TTC Phe 40 TAC TYT	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr	ol Allio GTG Val GTT Val CAG Gln	GGG GLY GGC GLY GGC GLY GGTA	AAC Asn GGG Gly 45 TTT Phe	TGG Trp 30 GGG Gly ACA Thr	GCT Ala ACC Thr CCC Pro	TTC Phe GGC	97 - 145
30 35	GTC Val GTC Val GTG Val TTT Phe 65	TTG Leu  ACA Thr 50 TCA Ser	GCG Ala GTT Val 35 GGG Gly CGA Arg	GGC Gly 20 GTG Val GGG Gly GGG GGG GG	CTT Leu ATG Met AAG Lys CCG Pro	ta Va  5 GCC Ala CTG Leu GTA Val TCT Ser 70	TAC Tyr CTC Leu GCC Ala 55 CAG Gln	TAT TYT TTC Phe 40 TAC Tyr	TCC Ser 25 GCC Ala ACC Thr	ATG Met GGT Gly ACC Thr	GTT Val  CAG GIn  CTT Leu 75	GGG Gly GAC Asp GGC Gly 60 GTA Val	AAC Asn GGG Gly 45 TTT Phe AAC Asn	TGG Trp 30 GGG Gly ACA Thr	GCT Ala ACC Thr CCC Pro	TTC Phe GGC Gly 80	97 - 145
30 35	GTC Val GTC Val GTG Val TTT Phe 65 AGC	TTG TCA Ser	GCG Ala GTT Val 35 GGG Gly CGA Arg	GGC Gly 20 GTG Val GGG Gly GGG Gly	CTT Leu ATG Met AAG Lys CCG Pro	GCC Ala CTG Leu GTA Val TCT Ser 70 AGG	TAC Tyr CTC Leu GCC Ala 55 CAG Gln ACT	TAT TYT TTC Phe 40 TAC TYT AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC	ATG Met GGT Gly ACC Thr CAA Gln	GTT Val CAG Gln CTT Leu 75	GGG Gly GGC Gly GTA Val	AAC AS GGG Gly 45 TTT Phe AAC AS GAC	TGG Trp 30 GGG Gly ACA Thr ACT Thr	GCT Ala ACC Thr CCC Pro AAC ASn	TTC Phe GGC Gly 80 AAC	97 - 145
30 35	GTC Val GTC Val GTG Val TTT Phe 65	TTG TCA Ser	GCG Ala GTT Val 35 GGG Gly CGA Arg	GGC Gly 20 GTG Val GGG Gly GGG Gly	CTT Leu ATG Met AAG Lys CCG Pro AAT	GCC Ala CTG Leu GTA Val TCT Ser 70 AGG	TAC TYT CTC Leu GCC Ala 55 CAG Gln ACT	TAT TYT TTC Phe 40 TAC TYT AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC	ATG Met GGT Gly ACC Thr CAA Gln	GTT Val CAG Gln CTT Leu 75	GGG Gly GGC Gly GTA Val	AAC AS GGG Gly 45 TTT Phe AAC AS GAC	TGG Trp 30 GGG Gly ACA Thr ACT Thr	GCT Ala ACC Thr CCC Pro AAC ASn	TTC Phe GGC Gly 80 AAC	97 145 193
30 35	GTC Val GTC Val GTG Val TTT Phe 65 AGC	TTG TCA Ser	GCG Ala GTT Val 35 GGG Gly CGA Arg	GGC Gly 20 GTG Val GGG Gly GGG Gly	CTT Leu ATG Met AAG Lys CCG Pro	GCC Ala CTG Leu GTA Val TCT Ser 70 AGG	TAC TYT CTC Leu GCC Ala 55 CAG Gln ACT	TAT TYT TTC Phe 40 TAC TYT AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC	ATG Met GGT Gly ACC Thr CAA Gln	GTT Val CAG Gln CTT Leu 75	GGG Gly GGC Gly GTA Val	AAC AS GGG Gly 45 TTT Phe AAC AS GAC	TGG Trp 30 GGG Gly ACA Thr ACT Thr	GCT Ala ACC Thr CCC Pro AAC ASn	TTC Phe GGC Gly 80 AAC	97 145 193
30 35	GTC Val GTC Val GTG Val TTT Phe 65 AGC	TTG TCA Ser	GCG Ala GTT Val 35 GGG Gly CGA Arg	GGC Gly 20 GTG Val GGG Gly GGG Gly	CTT Leu ATG Met AAG Lys CCG Pro AAT	GCC Ala CTG Leu GTA Val TCT Ser 70 AGG	TAC TYT CTC Leu GCC Ala 55 CAG Gln ACT	TAT TYT TTC Phe 40 TAC TYT AAA Lys	TCC Ser 25 GCC Ala ACC Thr ATC	ATG Met GGT Gly ACC Thr CAA Gln AAT	GTT Val CAG Gln CTT Leu 75	GGG Gly GGC Gly GTA Val	AAC AS GGG Gly 45 TTT Phe AAC AS GAC	TGG Trp 30 GGG Gly ACA Thr ACT Thr	GCT Ala ACC Thr CCC Pro AAC ASn CTT	TTC Phe GGC Gly 80 AAC	97 145 193

	ACC	GGG	TTC	CTT	GCC	GCG	CTG	TTC	TAC	ACC	CAC	AGC	TTC	AAC	GCG	TCC	337
	Thr	Gly	Phe	Leu	Ala	Ala	Leu	Phe	Tyr	Thr	His	Ser	Phe	Asn	Ala	Ser	
-				100					105					110			
5	GGA	TGT	CCG	GAG	CGT	ATG	GCC	GGT	TGC	CGC	CCC	ATT	GAC	GAG	TTC	GCT	385
	Gly	Cys	Pro	Glu	Arg	Met	Ala	Gly	Cys	Arg	Pro	Ile	Asp	Glu	Phe	Ala	
			115					120					125				
10	CAG	GGG	TGG	GGT	CCC	ATC	ACT	CAT	GTT	GTG	CCT	AAC	ATC	TCG	GAC	CAG	433
	Gln	Gly	Trp	Gly	Pro	Ile	Thr	His	Val	Val	Pro	Asn	Ile	Ser	Asp	Gln	
		130					135					140		•			
	AGG	CCC	TAT	TGC	TGG	CAC	TAC	GCG	CCT	CGA	CCG	TGT	GGT	ATC	GTA	CCC	481
15	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Ala	Pro	Arg	Pro	Cys	Gly	Ile	Val	Pro	
	145					150					155					160	
						GGT											529
	Ala	Ser	Gln	Val	Сув	Gly	Pro	Val	Tyr	Сув	Phe	Thr	Pro	Ser	Pro	Val	
20					165					170					175		
						GAT											577
	Val	Val	Gly		Thr	Asp	Arg	Phe		Ala	Pro	Thr	Tyr	Asn	Trp	Gly	
				180					185					190			
25						GTG											625
	Asn	Asn		Thr	Asp	Val	Leu		Leu	Asn	Asn	Thr		Pro	Pro	Gln	
			195					200					205				
						TGT											673
30	GTĀ		Trp	Phe	GIY	Cys		Trp	Met	Asn	Gly		Gly	Phe	Thr	Lys	
	100	210	~~~				215					220					
						CCG											721
35	225	cys	GIĀ	GIĀ	Pro	Pro	Cys	Asn	TTE	GIY	_	Val	GIY	Asn	Asn		
		x Cm	mcc	000	200	230	maa	mma	000		235	000	~~~			240	
						GAC											769
	Leu	1111	Cys	FIG	245	Asp	Cys	Pne	Arg		HIS	Pro	GIU	ALA		TYT	
40	מרמ	מממ	ጥርጥ	CCT		ccc	CCM	mcc	mmc	250	CCM	200	mca	OM N	255	G > m	017
						GGG											817
		2,0	010	260	001	Gry	110	11p	265	1117	FIO	ALG	Cys	270	AGT	птэ	
	TAC	CCA	TAC		CTC	TGG	CAC	ጥልጥ		ጥርር	ልሮሞ	ርሞር	אמר		אככ	አመሮ	865
45						Trp											003
	-2-		275					280		~ <u>,</u> 3			285	THE	1111	T16	
	TTC	AAG		AGG	ATG	TAT	GTG		GGC	GTG	GAA	CAC		Cdudo	CAA	ርር <del></del> ም	913
								200	555		J	~	*****		JEE	CL	713

	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	Arq	Leu	Glu	Ala	
		290					295					300	-				
5	GCA	TGC	AAT	TGG	ACC	CGA	GGA	GAG	CGT	TGT	GAC	TTG	GAG	GAC	AGG	GAT	961
5											Asp						
	305					310				_	315				5	320	
	AGA	TCA	GAG	CTT	AGC	CCG	CTA	TTG	CTG	TCC	ACA	ACA	GAG	TGG	CAG		1009
											Thr						
10					325					330					335		
	CTG	CCC	TGT	TCC	TTC	ACC	ACC	CTG	CCG	GCT	CTG	TCC	ACT	GGT		ATT	1057
											Leu						
				340					345					350			
15	CAT	CTC	CAT	CAG	AAC	ATC	GTG	GAC	GTG	CAA	TAT	CTG	TAC	GGC	ATA	GGG	1105
											Tyr						
			355					360			-		365	_			
	TCG	GCG	GTT	GTC	TCC	TTC	GCA	ATC	AAA	TGG	GAA	TAT	ATT	CTG	TTG	CTT	1153
20											Glu						
		370					375		_	_		380					
	TTC	CTC	CTC	CTG	GCG	GAC	GCG	CGC	GTC	TGT	GCC	TGC	TTG	TGG	ATG	ATG	1201
											Ala						
25	385					390					395	_		-		400	
	CTG	CTG	ATA	GCC	CAC	GCC	GAC	GCC	ACC	TTA	GAG	AA					1236
	Leu	Leu	Ile	Ala	His	Ala	Asp	Ala	Thr	Leu	Glu		٠				
					405					410							
30																	

SEQ ID NO:32

SEQUENCE LENGTH: 1236 base pairs

SEQUENCE TYPE: nucleic acid

35 STRANDEDNESS: double TOPOLOGY: linear ANTI-SENSE: No ORIGINAL SOURCE

40 ORGANISM: Hepatitis C virus IMMEDIATE EXPERIMENTAL SOURCE

CLONE: N27MX24B-1

C CGG ATC CCA CAA GCC GTG GTG GAT ATG GTG GCA GGG GCC CAC TGG GGA 49
Arg Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly

50

GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG  Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys  20 25 30  GTC TTG GTT GTG ATG CTG CTC TTC GCC GGT GTT GAC GGG GGG ACC CAC  Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His  35 40 45	97 145 193 241
Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys 20 25 30  GTC TTG GTT GTG ATG CTG CTC TTC GCC GGT GTT GAC GGG GGG ACC CAC Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His 35 40 45	193
20 25 30  GTC TTG GTT GTG ATG CTG CTC TTC GCC GGT GTT GAC GGG GGG ACC CAC  Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His  35 40 45	193
Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly Thr His 35 40 45	193
35 40 45	
10	
GTG ACA GGG GGG AAG GTA GCC TAC ACC CAG GGC TTT ACA CCC TTC	241
Val Thr Gly Gly Lys Val Ala Tyr Thr Thr Gln Gly Phe Thr Pro Phe	241
50 55 60	241
TTT TCA CGA GGG CCG TCT CAG AAA ATC CAA CTT GTA AAC ACT AAC GGC	
Phe Ser Arg Gly Pro Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly	
65 70 75 80	
AGC TGG CAC ATC AAT AGG ACT GCC CTC AAT TGC AAT GAC TCC CTT AAC	289
Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn	
85. 90 95	
ACC GGG TTC CTT GCC GCG CTG TTC TAC ACC CAC AGC TTC AAC GCG TCC	337
Thr Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Ser Phe Asn Ala Ser	
100 105 110	
GGA TGT CCG GAG CGT ATG GCC GGT TGC CGC CCC ATT GAC GAG TTC GCT	385
Gly Cys Pro Glu Arg Met Ala Gly Cys Arg Pro Ile Asp Glu Phe Ala	
115 120 125	
CAG GGG TGG GGT CCC ATC ACT CAT GTT GTG CCT AAC ATC TCG GAC CAG	433
Gin Gly Tip Gly Plo Tie Thi His val val Pro Ash Tie Ser Asp Gin	
130 135 140 AGG CCC TAT TGC TGG CAC TAC GCG CCT CGA CCG TGT GGT ATC GTA CCC	
Arg Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro	481
35 145 150 155 160	
GCG TCG CAG GTG TGT GGT CCG GTG TAT TGC TTC ACC CCA AGC CCT GTT	529
Ala Ser Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val	323
165 170 175	
40 GTG GTG GGG ACG ACC GAT CGT TTC GGC GCC CCC ACG TAC AAC TGG GGA	577
Val Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr Tyr Asn Trp Gly	3,,,
180 185 190	
AAC AAT GAG ACG GAT GTG CTA CTC CTC AAC AAC ACA CGG CCG CAG	625
Asn Asn Glu Thr Asp Val Leu Leu Asn Asn Thr Arg Pro Pro Gln	<del>-</del>
195 200 205	

	GGG	AAC	TGG	TTT	GGC	TGT	ACA	TGG	ATG	AAT	GGC	ACT	GGG	TTC	ACA	AAG	673
	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met	Asn	Gly	Thr	Gly	Phe	Thr	Lys	
5		210					215					220					
Ū	ACG	TGC	GGG	GGC	CCC	CCG	TGC	AAC	ATC	GGG	GGG	GTC	GGC	AAC	AAT	ACC	721
	Thr	Сув	Gly	Gly	Pro	Pro	Cys	Asn	Ile	Gly	Gly	Val	Gly	Asn	Asn	Thr	
	225					230					235			•		240	
10	TTG	ACT	TGC	CCC	ACG	GAC	TGC	TTC	CGG	AAG	CAC	CCC	GAG	GCC	ACT	TAC	769
.0	Leu	Thr	Cys	Pro	Thr	qaA	Cys	Phe	Arg	Lys	His	Pro	Glu	Ala	Thr	Tyr	
					245					250					255		
	ACA	AAA	TGT	GGT	TCG	GGG	CCT	TGG	TTG	ACG	CCT	AGG	TGC	CTA	GTT	CAT	817
15	Thr	Lys	Сув	Gly	Ser	Gly	Pro	Trp	Leu	Thr	Pro	Arg	Cys	Leu	Val	His	
7.5				260					265					270			
			TAC														865
	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	Thr	Val	Asn	Phe	Thr	Ile	
20			275					280					285				
20			GTT														913
	Phe		Val	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	Arg	Leu	Glu	Ala	
		290					295					300					
25			AAT														961
20		Сув	Asn	Trp	Thr		Gly	Glu	Arg	Сув		Leu	Glu	Asp	Arg	Asp	
	305					310					315					320	
			GAG														1009
30	Arg	Ser	Glu	Leu		Pro	Leu	Leu	Leu		Thr	Thr	Glu	Trp		Val	
30	ama				325					330					335		
			TGT														1057
	ren	PIO	Cys		Phe	Thr	Thr	Leu		Ala	Leu	Ser	Thr		Leu	Ile	
35	C N M	ama	a.m	340					345					350			
33			CAT														1105
	пте	Ten	His	GIN	ASI	11e	vaı		vaı	GIN	ıyr	ren		GIY	Пе	GIY	
	mcc	ccc	355	Cma	maa			360					365				
40			GTT														1153
40	ser	370	Val	Val	ser	Pne	_	iie	гув	rrp	GIU		11e	Leu	Leu	Leu	
	መመረ		CMC	CITIC	000	a.a	375	000	ama	mam	000	380					
			CTC														1201
4E	385	neu	Leu	ren	WIG		ATG	Arg	val	cys		cys	rea	rrp	met		
45		CTC	אוווא	ccc	CAC	390	CAC	aac	200	mm r	395					400	
	CIG	CIG	ATA	GCC	CAC	GCC	GAC	GCC	ACC	TTA	GAG	AA					1236

### Leu Leu Ile Ala His Ala Asp Ala Thr Leu Glu 405 410

5 10	SEQ ID SEQUENT SEQUENT STRAND TOPOLOG ANTI-S ORIGIN ORGANI IMMEDIA	CE L CE T EDNE GY: ENSE AL S SM: ATE	ENGT YPE: SS: line : No OURC: Hepa EXPE	nuc doub ar E titi:	leic le s C	aci	d	s									
	<b></b>																
20														GAG			48
		ATG	Trp	Leu		Met	Met	Leu	Leu		Ala	Gln	Ala	Glu		Ala	
	1 mmc	CNC	220	000	5	ama	ama			10					15	_	
														GGA			96
25	rea	GIU	Asn	ьец 20	vai	vaı	Leu	Asn		Ala	Ser	Met	Ala	Gly	Ala	His	
	ccr	ልጥሮ	ርሞር		mmæ	C Mm	CITIC	mmc	25 mmc	man	000	000	maa	30 TAC			
														TAC			144
	027		35	OCT	rne	Dea	AGI	40	rite	Сув	ATG	ALG	45	TYE	TTE	гĀ2	
30	GGC	AGG		GTC	CCY	GGG	GCG		ТАУ	GCT	עיזיכי	ሞልሞ		GTA	ሞርር	CCC	192
														Val			192
	•	50				2	55		-1-			60	1	•••	P	110	
	CTG	CTC	CTG	CTC	TTG	MTG	GCG	CTA	ccs	SCA	CGG	GCG	TAC	GCC	ATG	GAC	240
35														Ala			
	65					70					75		-			80	
	CGG	GAS	ATG	GCT	GCA	TCG	TGC	GGA	GGC	GCG	GTT	TTT	GTA	GGT	CTG	GTA	288
	Arg	Xae	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	Leu	Val	
40					85					90					95		
	CTC	YTG	ACC	TTG	TCA	CCA	TAC	TAC	AAA	GTG	TTC	CTC	GCT	ARG	CTC	ATA	336
	Leu	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Val	Phe	Leu	Ala	Xaf	Leu	Ile	
				100					105					110			
45	TGG	TGG	TTR	CAA	TAT	CTC	ATC	ACC	AGR	GCC	GAG	GCG	CAC	YTG	CAA	GTG	384
	Trp	Trp	Leu	Gln	Tyr	Leu	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	Gln	Val	

50

			115					120					125				
	TGG	ATY	CCC	CCY	CTY	AAC	GTY	CGG	GGR	GGC	CGC	GAY	GCC	ATC	ATC	CTY	432
_											Arg						
5		130					135					140					
	CTC	ACR	TGT	GCG	GTC	CAY	CCR	GAG	CTR	ATY	TTT	GAC	ATC	ACC	AAR	CTY	480
	Leu	Tre	Сув	Ala	Val	His	Pro	Glu	Leu	Ile	Phe	Asp	Ile	Thr	Lys	Leu	
10	145					150					155					160	
70											CTC						528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Xag	Xah	Thr	
					165					170					175		
15											CTC						576
••	Xai	Хај	Pro		Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Xak	Сув	Met	
				180					185					190			
											GTY						624
20	Leu	Val		Lys	Xal	Ala	Gly		His	Tyr	Val	Gln		Ala	Xam	Xan	
	AAC	CENC	195	CCB	CmC	202	COM	200					205				
											TAT						672
	272	210	AIG	AIG	Beu	1111	215	THE	TAT	Ado	Tyr	220	HIS	Leu	xaq	Pro	
25	CTG		SAY	TGG	GCC	CAY		GGC	СТА	CGB	GAC		GCG	CTD	CCP	CTW	720
											Asp						120
	225					230		1		9	235			<b>V</b> 4.1	72.14	240	
	GAG	ccc	GTT	GYC	TTC	TCT	GAY	ATG	GAG	ACY	AAG	ATC	ATC	ACC	TGG		768
30											Lys						
					245					250	-			•	255		
	GCA	GAC	ACY	GCG	GCG	TGT	GGG	GAC	ATC	ATT	TTG	GGC	CTA	CCW	GTC	TCC	816
	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro	Val	Ser	
35				260					265					270			
	GCC	CGG	AGG	GGC	AAC	GAG	ATA	CTC	CTC	GGA	CCG						849
	Ala	Arg	Arg	Gly	Asn	Glu	Ile	Leu	Leu	Gly	Pro						
			275					280									
40	Y:	C or	T		R:	A or	G		M	: A	or C	:		ĸ	: G	or T	
	s:	G or	C C		W:	A or	T		Н	: A	or C	or	T	В	: G	or T	or C
		: Al						ıb:						Xac	: Me	t or	Leu
		: Al						e :						Xaf	: Ly	s or	Arg
45		: G1	-					uh:						Xai	: G1	n or	Arg
	xaj	: Me	et or	. Val	L		Xa	k :	Met	or A	Ala			Xal	: Al	a or	Val

	Xam : I	eu o	r Ph	е		X	an :	Met	or	Val			Xao	: V	al or	Ile
	Xap : A	sp o	r Va	1		X	aq:	Thr	or	Ala			Xar	: A	sp or	His
5	Xas : A	la o	r Va	1												
5																
	SEQ ID NO:	34														
	SEQUENCE L	ENGT	H: 5	24 b	ase ;	pair	s									
	SEQUENCE T	YPE:	nuc	leic	aci	đ										
10	STRANDEDNE	SS:	doub.	le												
	TOPOLOGY:	line	ar													
	ANTI-SENSE	: No											•			
	ORIGINAL S	OURC	E													
15	ORGANISM:	Hepa	titi	s C ·	viru	5										
	IMMEDIATE	EXPE	RIME	NTAL	sou	RCE										
	CLONE: 026															
20	ATC ACG	TGG	GGG	GCA	GAG	ACG	GCG	GCG	TGT	GGG	GAC	ATC	ATC	TCG	GGT	48
	Ile Thr	Trp	Gly	Ala	Glu	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Ser	Gly	
	. 1			5					10					15		
	CTA CCC	GTT	TCC	GCC	CGA	AGG	GGG	ARG	GAG	CTG	CTT	TTG	GGR	CCG	GCC	96
25	Leu Pro	Val	Ser	Ala	Arg	Arg	Gly	Xaa	Glu	Leu	Leu	Leu	Gly	Pro	Ala	
			20					25					30			
	GAT AGT	TTT	GAC	GGG	CAG	GGG	TGG	CGA	CTC	CTT	GCG	CCT	ATC	ACG	GCC	144
	Asp Ser	Phe	Asp	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	
30		35					40					45				
	TAC TCC	CAG	CAR	ACG	CGG	GGC	CTG	CTT	GGT	TGC	ATC	ATC	ACY	AGC	CTT	192
	Tyr Ser	Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Сув	Ile	Ile	Tre	Ser	Leu	
	50					55					60					
35	ACG GGC															240
	Thr Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	
	65				70					75					80	
	ACC GCA															288
40	Thr Ala	Thr	Gln	Ser	Phe	Leu	Ala	Thr	Cys	Xab	Asn	Gly	Val	Cys	Trp	
				85					90					95		
	ACT GTT															336
	Thr Val	Phe		Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	
45			100					105					110			
	CCA ATC	ACC	CAA	ATG	TAC	ACC	TAA	GTR	GAT	CAG	GAC	CTC	GTC	GGY	TGG	384

	Pro	Ile	Thr		Met	Tyr	Thr	: Asr 120		. Asp	Gln	Asp			G13	Trp	
	TCG	GCC			SGG	GCG	ССТ			י אריא	CCT		125		000	AGC	
5	Ser	Ala	Pro	Pro	Xac	Ala	Ara	Ser	Ten	Thr	Dro	Cre	ACC	TGC	GGC	Ser	432
		130					135		. 160	1111	· FIC	140		cys	GTÄ	Ser	
	TCG	GAC	CTT	TAT	TTG	GTC	ACG	AGF	CAT	GCI	GAT			CCG	GTG	CAC	480
10																His	100
70	145					150					155		•			160	
														CCC	AT		524
	Arg	Arg	Gly	Asp	Ser	Arg	Gly	Ser	Leu	Leu	Ser	Pro	Gly	Pro			
15					165					170							
		Co				A o			M	: A	or	С		K	: G	or T	
		Go		_		A o					or		T	В	: G	or T	or C
	хаа	: A	rg o	r Ly	S		X	ab :	Val	or	Ile			Xac	: G	ly or	Arg
20	SEQ ID	NO.	35														
	SEQUEN			н. O	21 h	960		_									
	SEQUEN						_	5									
	STRAND					401	•										
25	TOPOLO																
	ANTI-S	ENSE	: No														
	ORIGIN	AL S	OURC	Е													
	ORGANIS	SM:	Hepa	titis	3 C 1	/irus	3										
30	IMMEDIA	ATE :	EXPE	RIME	NTAL	SOUR	RCE				•						
	CLONE:	N23															
05														TCG			48
35		Leu	Ser	Pro		Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	Gly	
	1 CCC	CITIC	O.T.C	maa	5					10					15		
														CGG			96
40	FIO	rea	Add	20	Pro	Ser	Gly	Xab		Val	Gly	Ile	Phe	Arg	Ala	Ala	
	CTC	ጥርረ	»CC		CCC	cmm	CCC		25					30			
														CCC			144
		-1-	35	-mg	GLY	AGI	ATG	40	ATA	vaı	Asp	Pne		Pro	Val	Glu	
45	тст	ATG		ACC	ACY	ATG	CGG		CCG	ርጥር	ጥጥር	RCG	45 CAT	AAC	mæ s	3.00	100
														Asn			192
							,				· · ·		P	11	JUL	****	

		50					55					60					
	ccc	CCG	GCC	GTA	CCG	CAG	WCA	TTC	CAA	GTG	GCC		СТА	CAC	GCT	CCC	240
											Ala						
5	65					70					75			-		80	
	ACT	GGC	AGC	GGC	AAA	AGC	ACC	ARG	GTG	CCG	GCT	GCG	TAT	GCG	GCC	CAA	288
											Ala						
					85					90			-		95		
10	GGG	TAC	AAG	GTA	CTC	GTC	CTG	AAC	CCG	TCC	GTT	GCT	GCC	ACT	TTG	GGC	336
	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	
				100					105					110			
	TTT	GGG	GCG	TAY	ATG	TCC	AAG	GCA	CAT	GGT	GTT	GAC	CCT	AAC	ATC	AGA	384
15	Phe	Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Val	Asp	Pro	Asn	Ile	Arg	
			115					120					125				
	ACT	GGG	GTG	AGG	ACC	ATC	ACC	ACG	GGC	GCT	ccc	RTC	ACG	TAC	TCC	ACC	432
	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro	Xaf	Thr	Tyr	Ser	Thr	
20		130					135					140					
											TCT						480
		Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	
	145					150					155			•		160	
25											GAC						528
	Ile	Ile	Ile	Сув		Glu	Суѕ	His	Ser	Thr	Asp	Ser	Thr	Ser	Ile	Leu	
	000	<b>&gt;</b> ~~			165					170					175		
••											ACG						576
30	GIĀ	TTE	GIY		Val	Leu	Asp	GIn		Glu	Thr	Ala	Gly		Arg	Leu	* •
	cmc	CITIC	CMC	180	100	com			185					190			
											TCG						624
35	Val	Val	195	Ата	THE	ATG	THE		Pro	GIĀ	Ser	VAI		Val	Pro	His	
33	_С СФ	ልልሞ		GNG	CNC	CTC	CCC	200	maa	220	ACT	<i>aa</i> ,	205	<b>&gt;</b> ma	000	mma	620
											Thr						672
		210	110	GLU	Gru	VUI	215	nea	ser	Well	THE	220	GIU	iie	Pro	Pne	
40	ТАТ		AAG	GCC	ልጥሮ	CCC	_	GAG	CCC	እጥሮ	AAG		CCC	»cc	CAM	CITIC	720
70											Lys						720
	225	~-2	-,-			230	u	OLU	11.4.4	116	235	GIY	GIY	ALG	UTP	240	
		TTC	TGC	CAT	TCC		AAG	AAA	TGT	GAC	GAG	ርጥር	<b>ር</b> ርጥ	GCG	מממ		768
45											Glu						, 00
	_		<b>-</b>		245	-4 -			-4-	250	<b></b>				255		
															~~		

	TCG GCC	CTC GGA	GTC AA	Y GCT	GTA	GCA	TAY	TAC	CGG	GGT	CTT	GAT	GTG	816
	Ser Ala													
5		260				265					270			
•	TCC RTC	ATA CCG	ACA AG	C GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACW	GAC	GCT	864
	Ser Xah		Thr Se	r Gly	Asp	Val	Val	Val	۷al	Ala	Thr	Asp	Ala	
	· <b></b> _	275			280					285				
10	CTA ATG													912
	Leu Met	Thr Gly	Tyr Th		Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	
	290 ACA TGT	CMC		295					300					
														921
15	Thr Cys 305	val												
	Y : C or	- ф	R : A (	~ C			<b>v</b> .						_	_
	S : G or	_	W : A				M:						Gor	
	С			,			н:	A OI	C	T		в:	G or	ror
20	Xaa : Le	u or Pro	<b>)</b>	Xá	ab:	His	or A	ra				Yac	: Thr	
	Ala							-9				Auc	• 1111	OI
	Xad : Se	r or Thr	<u>-</u>	Xε	ie :	Lys	or A	urg				Xaf	: Ile	or
	Val					_		_						
25	Xag : Th	r or Ile	•	Χa	h:	Val	or I	le						
	SEQ ID NO:3													
	SEQUENCE LE			-	3									
30	SEQUENCE TY			.d										
	STRANDEDNES		.e											
	TOPOLOGY: 1. ANTI-SENSE:													
35	ORIGINAL SO													
39	ORGANISM: He			_										
	IMMEDIATE EX													
	CLONE: N16	al Diving	ITALI SUL	RCE										
40														
	GGC TAT	ACC GGC	GAC TTO	GAC	TCA (	GTYG	ልጥር ሳ	מארי י	ጥርሶ	አልሮ	ልሮኣ	τη <b>ζ</b> τα	CTC	40
	Gly Tyr													48
	1	-	5	<b>F</b>			10		cy o	non.	-11L	15	AGT	
45	ACC CAA	ACA GTC	-	AGC	TTG (	GAC		ACT '	TTC	ACC	АТУ		ACC	96
	Thr Gln													20
												GIU	Thr	

				20					25					30			
	ACG	ACC	GTA	CCC	CAA	GAT	GCG	GTG	TCG	CGC	TCG	CAG	CGG		GGC	AGG	144
		Thr															
5			35			_		40		_			45	5	1	5	
	ACT	GGT	AGG	GGC	AGR	GGG	GGC	ATA	TAC	AGG	TTT	GTA	ACT	CCA	GGG	GAA	192
	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
	•	50					55					60			_		
10	CGG	CCC	TCA	GGC	ATG	TTC	GAT	TCT	TCG	GTC	CTG	TGT	GAA	TGT	TAT	GAC	240
	Arg	Pro	Ser	Gly	Met	Phe	qaA	Ser	Ser	Val	Leu	Cys	Glu	Cys	Tyr	Asp	
	65					70					75					80	
	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACG	YCC	GCC	GAG	ACC	TCG	GTT	AGG	288
15	Ala	Gly	CAa	Ala	Trp	Tyr	Glu	Leu	Thr	Xaa	Ala	Glu	Thr	Ser	Val	Arg	
					85					90					95		
		CGG															336
	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys	Gln	Asp	His	
20				100					105					110			
		GAG															384
	Leu	Glu		Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
			115					120					125				
25		TTC															432
	His	Phe	Leu	Ser	Gln	Thr		Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	
	~~~	130					135					140					
		GCA															480
30		Ala	Tyr	GIn	Ala		Val	Сув	Ala	Arg		Lys	Ala	Pro	Pro	Pro	
	145	maa				150					155					160	
		TGG															528
	ser	Trp	Asp	GIN		Trp	ràs	Cys	Leu		Arg	Leu	Lys	Pro		Leu	
35	CNC	ccc	CCN	200	165	ome	mma	773 M		170					175		
		GGG															576
	UTP	Gly	PIO		PIO	Leu	Leu	dsx		Leu	GIY	Ala	Val		Asn	Xac	
	cmm	DCC.	amy	180	a. a	222			185					190			
40		RCC														TC	623
	VQI	Xad		THE	HIS	Pro	ITE		гĀ8	туг	IIe	Met		Сув	Met		
	v .	C or	195		n -	R	. ~	200		. .			205		_	_	
					R:											or T	
45													T				or C
	Add	: Pr	.0 01	. sei			БХ	: aı	ıyr	or E	ııs			Xac	: : G	ilu o	r Lys

Xad : Thr or Ala

	SEQ ID NO:37
5	SEQUENCE LENGTH: 623 base pairs
	SEQUENCE TYPE: nucleic acid
	STRANDEDNESS: double
	TOPOLOGY: linear
10	ANTI-SENSE: No
	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
	IMMEDIATE EXPERIMENTAL SOURCE
15	CLONE: U16-4

	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCG	GTG	ATC	GAC	TGT	AAT	ACA	TGT	GTC	48
	Gly	Tyr	Thr	Gly	Asp	Phe	qaA	Ser	Val	Ile	Asp	Cys	Asn	Thr	Сув	Val	
20	1				5					10					15		
	ATC	CAG	ACA	GTC	GAC	TTC	AGC	TTG	GAC	CCC	ACC	TTC	ACC	ATC	GAG	ACG	96
	Ile	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
				20					25					30			
25	ACT	ACC	GTG	CCC	CAA	GAC	GCG	GTG	TCA	CGC	TCG	CAA	CGG	CGA	GGC	AGG	144
	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
			35					40					45				
	ACT	GGC	AGG	GGC	AGG	CAA	GGC	ATT	TAC	AGG	TTT	GTG	ACT	CCA	GGA	GAA	192
30	Thr	${\tt Gly}$	Arg	Gly	Arg	Gln	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
		50					55					60					
	CGG	CCC	TCG	GGC	ATG	TTC	GAT	TCC	TCG	GTC	CTG	TGC	GAG	TGC	TAT	GAC	240
	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Сув	Glu	Cys	Tyr	Asp	
35	65					70					75					80	
	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	CCG	CCC	GCC	GAG	ACC	ACG	GTC	AGG	288
	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Pro	Pro	Ala	Glu	Thr	Thr	Val	Arg	
					85					90					95		
40	TTG	CGG	GCT	TAC	CTG	AAC	ACC	CCA	GGG	CTG	CCC	GTC	TGC	CAG	GAC	CAT	336
	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys	Gln	Asp	His	
				100					105					110			
	CTG	GAG	TTC	TGG	GAG	AGC	GTC	TTC	ACA	GGC	CTC	ACC	CAC	ATA	GAT	GCC	384
45							Val										
			115					120		_			125		-		

50

	CAC	ттс	TTG	TCC	CAG	ACC	AAG	CAA	GCA	GGA	GAC	ልልጥ	ርሞር	CCT	መልሮ	CTC	432
			Leu														432
		130					135					140			-1-		
5	GTA	GCG	TAC	CAA	GCA	ACA	GTG	TGC	GCT	AGA	GCT	CAG	GCT	CCA	CCT	CCA	480
	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Gln	Ala	Pro	Pro	Pro	
	145					150					155					160	
10	TCA	TGG	GAT	CAA	ATG	TGG	AAG	TGT	CTC	ATA	CGG	CTA	AAA	CCT	ACA	CTA	528
10	Ser	Trp	Asp	Gln	Met	Trp	Lys	Суз	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	
					165					170					175		
			CCA														576
15	Arg	Gly	Pro		Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	
				180					185					190			
			CTC													TC	623
	Val	Asn	Leu	Thr	His	Pro	Val		Lys	Tyr	Ile	Met		Сув	Met		
20			195					200					205				
	SEQ ID	NO•	38														
	SEQUEN			ı. 6	IR ha	960 1	na i ro	•									
	SEQUEN					_		•									
25	STRAND	-				4011	•										
	TOPOLO	GY:	linea	ar													
	ANTI-S	ENSE:	: No														
	ORIGIN	AL S	OURCI	E													
30	ORGANI	SM: I	Hepat	titia	s C 1	virus	3										
	IMMEDI.	ATE 1	EXPE	RIME	MTAL	SOU	RCE										
	CLONE:	N13.	-1														
35	GCG	GATC(CAAA	50
					ı Thi	r His			Ala	a His	Phe	e Let	ı Seı	Glr	1 Thi	Lys	
	a. a]									10					
40			GGA														98
40		AIA	Gly	Asp	Asn		Pro	Tyr	Leu	Val		Tyr	Gln	Ala	Thr		
	15 TCC	cca	200	200		20	002	~~			25					30	
			AGG														146
45	cys	ALG	Arg	wra	ъув 35	WIG	PIO	PEO	PEO	Ser 40	тгр	Asp	GIN	met	_	тÀв	
40	ፐርጥ	CTC	ATA	CGG		AAG	ССФ	ACG	ሮሞል		GGG	CCA	ACG	CCC	45	መጥር፤	194
						-440		-100	OTM	CAC	333	CUA	AUG		CIG	110	174

	Cys	reu	TTE	Arg	Leu	Lys	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu		
				50					55				•	60				
	TAT	' AGG	TTA	GGA	GCC	GTT	CAG	AAC	GAG	GTT	ACC	CTC	ACA	CAC	CCC	ATA	242	
5	Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	Val	Thr	Leu	Thr	His	Pro	Ile		
			65					70					75					
	ACC	AAG	TTC	ATC	ATG	GCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	290	
	Thr	Lys	Phe	Ile	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr		
10		80					85					90						
	AGC	ACT	TGG	GTG	CTG	GTA	GGC	GGG	GTC	CTC	GCG	GCT	CTG	GCC	GCG	TAC	338	
	Ser	Thr	Trp	Val	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tvr	555	
	95					100					105					110		
15	TGC	CTG	ACA	ACG	GGC	AGC	GTG	GTC	ATT	GTG	GGC	AGG	ATC	GTC	TTG		386	
	Cys	Leu	Thr	Thr	Gly	Ser	Val	Val	Ile	Val	Gly	Arq	Ile	Val	Leu	Ser	300	
					115					120	-	_			125			
	GGG	AGG	CCG	GTT	GTT	ATT	CCC	GAC	AGG	GAA	GTT	CTC	TAC	CAA		արոր	434	
20	Gly	Arg	Pro	Val	Val	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	101	
				130					135				•	140				
	GAT	GAA	ATG	GAA	GAG	TGC	GCC	TCG	CAC	CTC	CCT	TAC	ATC	GAA	CAA	GGA	482	
	Asp	Glu	Met	Glu	Glu	Сув	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Glv	102	
25			145					150				_	155			1		
	ATG	CAG	CTC	GCC	GAG	CAA	TTC	AAG	CAG	AAG	GCG	CTC	GGT	TTG	CTG	CAA	530	
	Met	Gln	Leu	Ala	Glu	Gln	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	330	
		160					165					170	-					
30	ACA	GCC	ACC	AAG	CAA	GCG	GAG	GCT	GCT	GCT	CCC	GTG	GTG	GAG	TCC	AAG	578	
•	Thr	Ala	Thr	Lys	Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Ivs	3,73	
	175					180					185					190		
	TGG	CGA	GCC	CTT	GAG	ACC	TTC	TGG	GCG	AAG	CA G	GATO	CGC			-20	618	
35	Trp	Arg	Ala	Leu	Glu	Thr	Phe	Trp	Ala	Lys							210	
J.J					195			-		200								

SEQ ID NO:39

SEQUENCE LENGTH: 969 base pairs
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ANTI-SENSE: No

45 ORIGINAL SOURCE

50

ORGANISM: Hepatitis C virus IMMEDIATE EXPERIMENTAL SOURCE CLONE: N15-1

5

	GCG	GATC	CT C	CA C	CT C	CA T	CG T	GG G	AT C	AA A	TG T	GG A	AG I	GT C	TC A	TA CGG	5 5 1
			P	ro P	ro P	ro S	er T	rp A	sp G	ln M	et T	rp L	ys C	ys L	eu I	le Arç	Ī
40	•			1				5					10				
10																GGA	99
		Lys	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	
	15					20					25					30	
15									ACA								147
75	Ala	Val	Gln	Asn		Val	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	Phe	Ile	
					35					40					45		
									GAG								195
20	Met	Ala	Сув		Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	
20				50					55					60			
									CTG								243
	Leu	Val		Gly	Val	Leu	Ala		Leu	Ala	Ala	Tyr	_		Thr	Thr	
05			65					70		•			75				
25									ATC								291
	GIĀ		Val	Val	Ile	Val		Arg	Ile	Ile	Leu			Arg	Pro	Ala	
		80					85					90					
									TAC								33 9
30		Ile	Pro	qaA	Arg		Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	
	95					100					105					110	
									ATC								38 7
	GIU	Cys	Ala	Ser		Leu	Pro	Tyr	Ile		Gln	Gly	Met	Gln	Leu	Ala	
35					115					120					125		
									GGT								435
	GIU	GIn	Pne		GIn	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	
	~~~			130					135					140			
40									GTG								483
	GIN	ATG		Ala	Ala	Ala	Pro		Val	Glu	Ser	Lys		Arg	Ala	Leu	
	a. a		145					150					155				
									TGG _								531
45	GIU		rue	TTP	ALA	ГЛS		Met	Trp	Asn	Phe		Ser	Gly	Ile	Gln	
		160					165					170					

50

	TAC	TTA	GCA	GGC	TTG	TCC	ACT	CTG	CCT	GGA	AAC	ccc	GCA	ATA	GCA	TCA	579
					Leu												
	175					180					185					190	
5	CTG	ATG	GCA	TTC	ACA	GCC	TCT	ATC	ACC	AGC	CCG	CTC	ACC	ACC	CAA	TAT	627
	Leu	Met	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	Gln	Tyr	
					195					200					205		
	ACC	CTC	CTG	TTT	AAC	ATC	TTG	GGG	GGA	TGG	GTG	GCC	GCC	CAA	CTC	GCC	675
10	Thr	Leu	Leu	Phe	Asn	Ile	Leu	Gly	Gly	Trp	Val	Ala	Ala	Gln	Leu	Ala	
				210					215					220			
					GCT												723
	Pro	Pro		Ala	Ala	Ser	Ala	Phe	Val	Gly	Ala	Gly	Ile	Ala	Gly	Ala	
15			225		•			230					235				
					ATA												771
	Ala		Gly	Ser	Ile	Gly	Leu	Gly	Lys	Val	Leu	Val	Asp	Ile	Leu	Ala	
		240					245					250					
20					GGG												81 <b>9</b>
		Tyr	Gly	Ala	Gly		Ala	Gly	Ala	Leu	Val	Ala	Phe	Lys	Val	Met	
	255					260					265					270	
					CCC												867
25	Ser	Gly	Asp	Met	Pro	Ser	Thr	Glu	qaA	Leu	Val	Asn	Leu	Leu	Pro	Ala	
					275					280					285		
					GGT												915
	ııe	ren	ser		Gly	Ala	Leu	Val		Gly	Val	Val	Суз		Ala	Ile	
30	CITIC	OCE	000	290	ama				295					300			
					GTG												963
	neu	Arg	305	HIS	Val	GIĀ	Pro		GIU	GLY	Ala	Val		Trp	Met	Asn	
	CGC	ריתיב	C AG	·cc				310					315				
35	Arg		C Ac	icc													974
	AL Y	320															
		320															
	CDO TD																

SEQ ID NO:40

SEQUENCE LENGTH: 1280 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
ANTI-SENSE: No

50

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

5 CLONE: MX25026

					TGG												48
40	Cys	Ala	Trp	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	
10	1				5					10					15		
	TTG	GAG	AAC	CTG	GTG	GTC	CTC	AAT	GCA	GCA	TCC	ATG	GCS	GGA	GCG	CAT	96
	Leu	Glu	Asn	Leu	Val	Val	Leu	Asn	Ala	Ala	Ser	Met	Ala	Gly	Ala	His	
				20					25					30			
15	GGC	ATC	CTC	TCT	TTC	CTT	GTG	TTC	TTC	TGT	GCC	GCC	TGG	TAC	ATC	AAA	144
	Gly	Ile	Leu	Ser	Phe	Leu	Val	Phe	Phe	Cys	Ala	Ala.	Trp	Tyr	Ile	Lys	
			35					40					45				
	GGC	AGG	CTG	GTC	CCY	GGG	GCG	RÇA	TAY	GCT	YTC	TAT	GGC	GTA	TGG	CCG	192
20	Gly	Arg	Leu	Val	Pro	Gly	Ala	Xaa	Tyr	Ala	Xab	Tyr	Gly	Val	Trp	Pro	
		50					55					60					
	CTG	CTC	CTG	CTC	TTG	MTG	GCG	CTA	ccs	SCA	CGG	GCG	TAC	GCC	ATG	GAC	.240
	Leu	Leu	Leu	Leu	Leu	Xac	Ala	Leu	Pro	Xad	Arg	Ala	Tyr	Ala	Met	Asp	
25	65					70					75					80	
	CGG	GAS	ATG	GCT	GCA	TCG	TGC	GGA	GGC	GCG	GTT	TTT	GTA	GGT	CTG	GTA	288
	Arg	Xae	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	Leu	Val	
					85					90					95		
30	CTC	YTG	ACC	TTG	TCA	CCA	TAC	TAC	AAA	GTG	TTC	CTC	GCT	ARG	CTC	ATA	336
	Leu	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Val	Phe	Leu	Ala	Xaf	Leu	Ile	
				100					105					110			
	TGG	TGG	TTR	CAA	TAT	CTC	ATC	ACC	AGR	GCC	GAG	GCG	CAC	YTG	CAA	GTG	384
35	Trp	Trp	Leu	Gln	Tyr	Leu	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	Gln	Val	
			115					120					125				
	TGG	ATY	CCC	CCY	CTY	AAC	GTY	CGG	GGR	GGC	CGC	GAY	GCC	ATC	ATC	CTY	432
	Trp	Ile	Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Ile	Ile	Leu	
40		130					135					140					
	CTC	ACR	TGT	GCG	GTC	CAY	CCR	GAG	CTR	ATY	TTT	GAC	ATC	ACC	AAR	CTY	480
	Leu	Thr	Cys	Ala	Val	His	Pro	Glu	Leu	Ile	Phe	Asp	Ile	Thr	Lys	Leu	
	145					150					155			-		160	
45	YTG	CTC	GCC	ATA	CTC	GGT	CCG	CTC	ATG	GTR	CTC	CAG	GCT	GSC	MTA	ACY	528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Xag	Xah	Thr	

50

MRA RTG   CCG   TAC   TTY   GTR   CGY   GCT   CAA   GGG   CTC   ATY   CGT   RYG   TGC   ATG						165	,				170	)				175		
Sai Xaj Pro Tyr Phe Val Arg Ala Gln Gly Leu Ile Arg Xak Cys Met		MRA	RTG	CCG	TAC	TTY	GTR	CGY	GCI	CAA			: ATY	CGT	RYG			576
180		Xai	Хаj	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Glv	Leu	ı Ile	Aro	Xak	Cva	Mot	370
Leu Val Arg Lys Xal Ala Gly Gly His Tyr Val Gln Met Ala Xam Xam	J							_						3			*****	
Leu Val Arg Lys Xal Ala Gly Gly His Tyr Val Gln Met Ala Xam Xam		TTR	GTG	CGG	AAA	GYC	GCY	GGR	GGT	CAT	TAT	GTY	CAR	ATG	GCY	YTY	RTG	624
195		Leu	Val	Arg	Lys	Xal	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Xam	Xan	
AAG CTG GCY GCR CTG ACA GGT ACG TAC RTT TAT GWC CAT CTT RCY CCA Lys Leu Ala Ala Leu Thr Gly Thr Tyr Xao Tyr Xap His Leu Xaq Pro 210	10	•																
Lys   Leu   Ala   Ala   Leu   Thr   Gly   Thr   Tyr   Xao   Tyr   Xap   His   Leu   Xaq   Pro   210   215   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220   220		AAG	CTG	GCY	GCR	CTG	ACA	GGT	ACG	TAC	RTT	TAT	GWC	CAT	CTT	RCY	CCA	672
210		Lys	Leu	Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Xao	Tyr	Xap	His	Leu	Xaq	Pro	
Leu Gln Xar Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val Ala Val   225   230   240   235   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240			210					215					220					
Leu Gln Xar Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val Ala Val   225   230   240   235   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240   240	15	CTG	CAG	SAY	TGG	GCC	CAY	GCG	GGC	CTA	CGR	GAC	CTT	GCG	GTR	GCR	GTW	720
GAG CCC GTT GYC TCT GAY ATG GAG ACY AAG ATG ACG ACG ACG ACG ACG ACG ACG ACG ACG AC		Leu	Gln	Xar	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	
Column																		
245		GAG	CCC	GTT	GYC	TTC	TCT	GAY	ATG	GAG	ACY	AAG	ATC	ATC	ACS	TGG	GGG	768
GCA GAS ACB GCG GCG GCG GAC ACC ACC ACC ACC ACC ACC	20	GIU	Pro	Val	Xas		Ser	Asp	Met	Glu	Thr	Lys	Ile	Ile	Thr	Trp	Gly	
Ala Xat Thr Ala Ala Cys Gly Asp Ile Ile Xau Gly Leu Pro Val Ser  260		CCA	<b>636</b>	<b>1</b>	000													
260 CGR AGG GGY ARS GAG MTR CTY YTS GGR CCG GCC GAT AGT TTT GAC 864 Ala Arg Arg Gly Xav Glu Xaw Leu Xax Gly Pro Ala Asp Ser Phe Asp  275																		816
GCC CGR AGG GGY ARS GAG MTR CTY YTS GGR CCG GCC GAT AGT TTT GAC Ala Arg Arg Gly Xav Glu Xaw Leu Xax Gly Pro Ala Asp Ser Phe Asp 275		VIG	Adl	THE		AIA	Cys	GLY	Asp		Ile	Xau	Gly	Leu		Val	Ser	
Ala Arg Arg Gly Xav Glu Xaw Leu Xax Gly Fro Ala Asp Ser Phe Asp  275	25	GCC	CCB	) ACC		ADC	CAC	Mmn	CONT									
GGG CAG GGG TGG CGA CTC CTT GCG CCT ATC ACG GCC TAC TCC CAG CAR 912  30 Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln 290																		864
GGG CAG GGG TGG CGA CTC CTT GCG CCT ATC ACG GCC TAC TCC CAG CAR 912  Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln C290			9		Q ₁	Adv	Gru	AGW		лах	GTÄ	PTO	ATA		Ser	Phe	Asp	
Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln 290 295 300  ACG CGG GGC CTG CTT GGT TGC ATC ATC ACY AGC CTT ACG GGC CGG GAT 960  Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp 35 305 310 315 320  AAR AAC CAG GTC GAG GGG GAG GTT CAA GTG GTC TCT ACC GCA ACA CAA 1008  Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln 325 330 335  40 TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056 Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His 340 345 345 350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  45 Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		GGG	CAG		TGG	CGA	ርሞሮ	Cdrds		CCM	a m.c	300	000		<b></b>	~~~		•••
290	30	Gly	Gln	Glv	Tro	Ara	Len	Len	Δla	Dro	TIO	Mb~	BT n	TAC	TCC	CAG	CAR	912
ACG CGG GGC CTG CTT GGT TGC ATC ATC ACY AGC CTT ACG GGC CGG GAT 960  Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp  35 305		-		4		9			ALC.	110	116	THE		TYE	ser	GIN	GIN	
Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp  305		ACG	CGG	GGC	CTG	CTT	GGT		ATC	ATC	ACY	AGC		ACG.	GGC	ccc	Cam	060
310  315  320  AAR AAC CAG GTC GAG GGG GAG GTT CAA GTG GTC TCT ACC GCA ACA CAA 1008  Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln  325  325  330  335  40  TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056  Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His  340  340  345  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  45  Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Glv	Ara	Acn	360
AAR AAC CAG GTC GAG GGG GAG GTT CAA GTG GTC TCT ACC GCA ACA CAA 1008  Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln  325 330 335  40 TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056  Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His  340 345 350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  45 Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln	35	305						-							7	<u>-</u>		
Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln  325 330 335  40 TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056  Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His  340 345 350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  45 Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		AAR	AAC	CAG	GTC	GAG	GGG	GAG	GTT	CAA	GTG	GTC	TCT	ACC	GCA	ACA		1008
325 330 335 40  TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056  Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His 340 345 350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  45 Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	Thr	Ala	Thr	Gln	2000
TCT TTC CTG GCG ACC TGY RTC AAC GGC GTK TGC TGG ACT GTT TTC CAC 1056  Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His  340  345  350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln																		
Ser Phe Leu Ala Thr Cys Xay Asn Gly Val Cys Trp Thr Val Phe His  340  345  350  GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104  Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln	40	TCT	TTC	CTG	GCG	ACC	TGY	RTC	AAC	GGC	GTK	TGC	TGG	ACT	GTT	TTC	CAC	1056
340 345 350 GGY GCC GGC TCG AAG ACC TTA GCC GGC CCA AAA GGC CCA ATC ACC CAA 1104 Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		Ser	Phe	Leu	Ala	Thr	Cys	Хау	Asn	Gly	Val	Cys	Trp	Thr	Val	Phe	His	
Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln													_					
Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln		GGY	GCC	GGC	TCG	AAG	ACC	TTA	GCC	GGC	CCA	AAA	GGC	CCA	ATC	ACC	CAA	1104
255	45	Gly .	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	Pro	Ile	Thr	Gln	

	ATG TAC ACC AAT GTR GAT CAG GAC CTC GTC GGY TGG TCG GCG CCC CCC 11	
	Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Ser Ala Pro Pro	52
	370 375 380	
5	SCC CCC CCM MCC MMC ACA COV MCC ACC ACC ACC ACC	00
	Xaz Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr	00
	385 390 395 400	
	MINIC CINC ACC ACC ACC CAN COM CAN COM	48
10	Leu Val Thr Arg His Ala Asp Val Ile Pro Val His Arg Arg Gly Asp	40
	405 410 415	
	NCC NCC CCC NCC CMC CMC MCC CCC CCC CCC	80
	Ser Arg Gly Ser Leu Leu Ser Pro Gly Pro	
15	420 425	
	Y: CorT R: AorG M: AorC K: GorT	
	S:GorC W:AorT H:AorCorT B:GorTor	С
	Xaa : Ala or Thr Xab : Phe or Leu Xac : Met or Le	u
20	Xad : Ala or Pro Xae : Glu or Asp Xaf : Lys or Ar	g
	Xag : Gly or Ala Xah : Leu or Ile Xai : Gln or Ar	g
	Xaj: Met or Val Xak: Met or Ala Xal: Ala or Va	1
	Xam : Leu or Phe Xan : Met or Val Xao : Val or Il	е
25	Xap : Asp or Val Xaq : Thr or Ala Xar : Asp or Hi	В
	Xas : Ala or Val Xat : Asp or Glu Xau : Leu or Se	r
	Xav : Asn or Arg or Lys Xaw : Ile or Leu Xax : Leu or Ph	е
00	Xay : Ile or Val Xaz : Gly or Arg	
30	GEO. TR. NO. 44	
	SEQ ID NO:41	
	SEQUENCE LENGTH: 1431 base pairs SEQUENCE TYPE: nucleic acid	
35	STRANDEDNESS: double	
33	TOPOLOGY: linear	
	ANTI-SENSE: No	
	ORIGINAL SOURCE	
40	ORGANISM: Hepatitis C virus	
<del></del> -	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: N16N15	
45	GGC TAT ACC GGC GAC TTC GAC TCA GTG ATC GAC TGC AAC ACA TGT GTC 4	
70	Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val	>
	7 -1 CAS AND THE MEN CAS WELL THE CAS ASI	

	1				5					10					15		
	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC			TTC	ACC	ATTY		ACG	96
	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Авр	Pro	Thr	Phe	Thr	Ile	Glu	Thr	30
5				20					25					30			
	ACG	ACC	GTA	ccc	CAA	GAT	GCG	GTG	TCG	CGC	TCG	CAG	CGG	CGA	GGC	AGG	144
												Gln					
	•		35					40					45	-	•		
10	ACT	GGT	AGG	GGC	AGR	GGG	GGC	ATA	TAC	AGG	TTT	GTA	ACT	CCA	GGG	GAA	192
	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
		50					55					60	•		_		
	CGG	CCC	TCA	GGC	ATG	TTC	GAT	TCT	TCG	GTC	CTG	TGT	GAA	TGT	TAT	GAC	240
15	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	Glu	Сув	Tyr	Asp	
	65					70					75					80	
												GAG					288
	Ala	Gly	Сув	Ala	Trp	Tyr	Glu	Leu	Thr	Xaa	Ala	Glu	Thr	Ser	Val	Arg	
20					85					90					95		
												GTC					336
	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Сув	Gln	Asp	His	
				100					105					110			
25												ACC					384
	Leu	Glu		Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
			115					120					125				
												AAC					432
30	His		Leu	Ser	Gln	Thr	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	
		130					135					140					
												AAG					480
		Ala	туг	GIn	Ala		Val	Cys	Ala	Arg		Lys	Ala	Pro	Pro	Pro	
35	145	mcc.	<b>63.</b> m	<b>03</b> n		150					155					160	
												CTG					528
	Ser	пр	Asp	GIN		ттр	гÀа	Cys	Leu		Arg	Leu	Lys	Pro		Leu	
	CAC	ccc	COX	3.00	165	<b>a</b> ma	mm.a			170					175		
40												GCC					. 576
	urs	GTĀ	PLO		Pro	тел	Leu	Xab		Leu	Gly	Ala,	Val		Asn	Xac	
	Cum	DCC.	ር ሙህ	180	as a	00**	3005	3.00	185					190			
												ATG					624
45	*41			THE	uTR	PIO	TTE		nĀs	xae	тте	Met		Cys	Met	Ser	
			195					200					205				

	GCT	GAC	СТА	GAG	GTC	GTC	ACT	AGC	ACT	TGG	GTG	CTG	GTA	GGC	GGG	GTC	672
					Val												
_		210					215			_		220		-	-		
5	CTC	GCG	GCT	CTG	GCC	GCG	TAC	TGC	CTG	ACA	ACG	GGC	AGC	GTG	GTC	ATT	720
	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	Gly	Ser	Val	Val	Ile	
	225					230					235					240	
10	GTG	GGC	AGG	ATC	ATC	TTG	TCC	GGG	AGG	CCG	GCC	GTT	ATT	CCC	GAC	AGG	768
10	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	Val	Ile	Pro	Asp	Arg	
					245					250					255		
	GAA	GTT	CTC	TAC	CAA	GAG	TTC	GAT	GAA	ATG	GAA	GAG	TGC	GCC	TCG	CAC	816
15	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	Сув	Ala	Ser	His	
15				260					265					270			
	CTC	CCT	TAC	ATC	GAA	CAA	GGA	ATG	CAG	CTC	GCC	GAG	CAA	TTC	AAG	CAG	864
	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu	Gln	Phe	Lys	Gln	
20			275					280					285				
20					TTG												912
	Lys		Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	ГÀв	Gln	Ala	Glu	Ala	Ala	
		290					295					300					
25					GAG												960
25		Pro	Val	Val	Glu		Lys	Trp	Arg	Ala		Glu	Thr	Phe	Trp	Ala	
	305	~~~				310					315		•			320	
					AAT												1008
30	гаг	HIS	Met	Trp	Asn	Phe	Ile	Ser	Gly		Gln	Tyr	Leu	Ala		Leu	
55	maa	<b>3</b> (7m)	ama		325					330					335		
					GGA												1056
	Ser	THE	ьеи	340	Gly	Asn	PIO	Ala		ATA	ser	Leu	Met		Phe	Thr	
35	GCC	መርጥ	ልሞሮ		AGC	ccc	CMC	200	345	C 2 2	mam	200	ama	350	mmm		
					Ser												1104
		001	355		UC.	110	Deu	360	1111	GIII	TYT	1111	365	ьeu	Pile	ASII	
	ATC	TTG		GGA	TGG	CTC	GCC		CAA	כיזיכי	GCC	מממ.		እርጥ	ccc	CCB	1152
40					Trp												1152
		370	~- <i>1</i>	1			375	1114	<b>011</b>	ДСС	ALU	380	110	Ser	ALG	MIG	
	TCA		TTC	GTG	GGC	GCC		АТА	CCT	GGC	GCG		ርጥጥ	ccc	<b>ACC</b>	እመአ	1200
					Gly												1200
45	385				1	390	1			1	395			- <u>1</u>		400	
		CTC	GGG	AAG	GTG		GTG	GAC	ATT	CTG		GCT	ጥፈጥ	GG A	GC A		1248
	-									J. G	500	201	****	JUA	JUA	333	1740

	Gly	Leu	Gly	Lys			Val	Asp	Ile			Gly	Tyr	Gly	Ala	Gly	
	CITIC	CC3	000	~~~	405					410					415		
5	Ua l	Ala	GGC	31-	CTC	GTG	GCC	TTT	AAG	GTC	ATG	AGC	GGT	GAC	ATC	ccc	1296
	AGI	wra	GIY			vai	Ala	Phe		Val	Met	Ser	Gly	Asp	Met	Pro	
	mcc	3.00	a.a	420					425					430			
	Com	Mb	GAG	GAC	CTG	GTC	AAC	TTA	CTC	CCC	GCC	ATC	CTC	TCT	CCI	GGT	1344
10	ser	Thr		Asp	Leu	Val	Asn		Leu	Pro	Ala	Ile	Leu	Ser	Pro	Gly	
	000		435					440					445				
	Sin Sin	CIG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	ATA	CTG	CĢT	CGG	CAI	GTG	1392
	Ala .	Leu	vaı	Val	Gly	Val		Cys	Ala	Ala	Ile	Leu	Arg	Arg	His	Val	
15		450					455					460					
	GGC (	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG				1431
	Gly :	Pro	GLY	Glu	GLy		Val	Gln	Trp	Met	Asn	Arg	Leu				
	465	<b>~</b>		-	_	470					475						
20	Y : (					A o	_				or (				: G	or T	
,	S : (			_		A o						or	T	В	: G	or T	or C
	Xaa .							ab:						Xac	: G	lu or	Lys
	Xad :	: Th	r or	: Ala	1		Xε	ie:	Tyr	or I	Phe			Xaf	: T	hr or	Ala
25	CPO ID 1	70 - A	•														
	SEQ ID 1																
	SEQUENCE						_	:8									
	SEQUENCI STRANDEI					acio	l										
30	TOPOLOGY																
	ANTI-SEN	-		-			•	•		٠							
	ORIGINAL			1													
	ORGANISM					ri ruc											
35	IMMEDIAT																
	CLONE: N					5001											
	CTG C	TG :	TCG	ccc	GGG	ccc	ATYC	ጥርV	ጥልሮ	ציוועב	AAC	CCV	mcc	mac		aam	4.0
40	Leu L	eu s	Ser	Pro	Glv	Pro	Ile	Ser	Tur	Ten	ANG Tara	C1	rcc	TCG	GGT	GGT	48
	1				5 5			DCI	-y-	10	nys	GIY	ser	ser		GIĀ	
	CCG C	TG (	CYT '	TGC	-	ጥሮር	ככר	י חמי	c mm		CCC	3.mo		~~~	15		
	Pro L	eu 2	Kaa (	Cvs	Pro	Ser	Clv	Yah '	us i	are j	GGC .	AIC	TTC	CGG	GCT	GCY	96
45				20			~- <i>y</i>	aab	25	AGT	GIĀ	TTE .	rne .		ALA	ALA	
	GTG T	GC 2	ACC (		GGG	GTT	GCG	AAG 4		උආත	ርአር ፡	mmar 4		30	a	016	
							JUG .	· Draw	JCG	JIK	GAC	TTT (	21.6	CCC	GTT	GAG	144

	Val	Cys	Thr	Arg	Gly	Val	Ala		Ala	Val	Asp	Phe		Pro	Val	Glu	
	mam	3 m/c		100				40					45				
5					ACY												192
	Ser	мес 50	GIU	Thr	Thr	Met		Ser	Pro	Val	Phe		Asp	Asn	Ser	Thr	
	000		000	am.	222	~~~	55					60					
					CCG												240
10	65	PIO	ATG	val	Pro		xad	Phe	Gin	Val		His	Leu	His	Ala		
		ccc	AGC	ccc		70	3.00	NDC.	ama	200	75					80	
					AAA												288
		GLY	261	GLY	Lys 85	Set	THE	Ade	vaı		Ата	ATA	Tyr	ALA		GIn	
15	ccc	ጥልሮ	AAC	Cmx		CMC	ama	220	000	90					95		
					CTC												336
	GLY	-71	шуа	100	Leu	VAI	reu	WRII		ser	vaı	ATA	ATA		Leu	GLY	
	սերեր	ccc	GCG		ATG	ምርር	AAG	CCA	105	ccm	cmm	<b>a</b> 2a	aam	110			204
20					Met												384
		011	115	+3+	Met	Del	цув	120	птв	GŢŽ	vaı	Asp		Asn	TTE	Arg	
	ACT	GGG		AGG	ACC	<b>አ</b> ጥር	ACC	_	ccc	CCM	000	Dma	125	m» a	maa		420
					Thr												432
25		130					135	****	GLY	ATO	FIO	140	THE	TYL	ser	THE	
	TAC		AAG	TTC	CTC	GCC		CCT	GGC	ጥረጥ	ጥርጥ		CCT	CCC	ጥአመ	CAC	400
					Leu												480
	145	-	•			150	<i>E</i>	1	01,	0,5	155	GLy	GLY	VIG	TÄT	160	
30	ATC	ATA	ATA	TGT	GAT	GAG	TGT	CAT	TCA	ACT	_	TCG	ልሮሞ	ጥርር	ልጥሮ		528
					Asp												220
				_	165		-			170					175		
	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG		ACG	GCT	GGA	GCG		СТТ	576
35					Val												0.0
				180			_		185				2	190	5		
	GTC	GTG	CTC	GCC	ACC	GCT	ACG	CCT	CCG	GGA	TCG	GTC	ACC	GTG	CCG	CAT	624
					Thr												
40			195					200		_			205				
	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA		ATC	CCC	TTC	672
					Glu												•
		210					215					220					
45	TAT	GGC	AAG	GCC	ATC	CCC	CTC	GAG	GCC	ATC	AAG	GGG	GGG	AGG	CAT	CTC	720
					Ile												
											-	-	-	_			

	225					230					235					240	
	AYT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	GCG	AAG	CTG	768
	Xag	Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	
5					245			_		250					255		
	TCG	GCC	CTC	GGA	GTC	AAY	GCT	GTA	GCA	TAY	TAC	CGG	GGT	CTT	GAT	GTG	816
	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	Val	
	•			260					265	<del>-</del> .	-	•	_	270	-		
10	TCC	RTC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACW	GAC	GCT	864
	Ser	Xah	Ile	Pro	Thr	Ser	Gly	Asp	Val	Val	Val	Val	Ala	Thr	Asp	Ala	
			275					280					285				
	CTA	ATG	ACG	GGC	TAT	ACC	GGY	GAC	TTY	GAC	TCR	GTG	ATC	GAC	TGY	AAC	912
15	Leu	Met	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Сув	Asn	
		290					295					300					
	ACA	TGT	GTC	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	CCT	ACT	TTC	ACC	960
	Thr	Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	
20	305					310					315					320	
	ATY	GAG	ACG	ACG	ACC	GTA	CCC	CAA	GAT	GCG	GTG	TCG	CGC	TCG	CAG	CGG	1008
	Ile	Glu	Thr	Thr	Thr	<b>V</b> al	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	
					325					330					335		
25	CGA	GGC	AGG	ACT	GGT	AGG	GGC	AGR	GGG	GGC	ATA	TAC	AGG	TTT	GTA	ACT	1056
	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Gly	${\tt Gly}$	Ile	Tyr	Arg	Phe	Val	Thr	
				340					345					350			
	CCA	GGG	GAA	CGG	CCC	TCA	GGC	ATG	TTC	GAT	TCT	TCG	GTC	CTG	TGT	GAA	1104
30	Pro	Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	Glu	
			355					360					365				
	TGT	TAT	GAC	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACG	YCC	GCC	GAG	ACC	1152
	Cys	_	Asp	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	Xai	Ala	Glu	Thr	
35		370					375					380					
														CCC			1200
		Val	Arg	Leu	Arg		Tyr	Leu	Asn	Thr		Gly	Leu	Pro	Val	-	
	385		_			390					395					400	
40			_											CTC			1248
	Gln	Asp	His	Leu		Phe	Trp	Glu	Ser		Phe	Thr	Gly	Leu		His	
					405					410					415		
														GAC			1296
<b>4</b> 5	He	Asp	Ala		Phe	Leu	Ser	Gln		Lys	Gln	Ala	Gly	Asp	Asn	Phe	
				420					425					430			

	CCC	TAC	CTG	GTA	GCA	TAC	CAG	GCT	ACA	GTG	TGC	GCC	AGG	GCC	AAG	GCT	1344
	Pro	Tyr	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Äla	Lys	Ala	
5			435					440					445				
Ū	CCA	CCT	CCA	TCG	TGG	GAT	CAR	ATG	TGG	AAG	TGT	CTC	ATA	CGG	CTG	AAG	1392
	Pro	Pro	Pro	Ser	Trp	Asp	Gln	Met	Trp	Lys	Сув	Leu	Ile	Arg	Leu	Lys	
		450					455					460					
10					GGG												1440
70	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Хај	Arg	Leu	Gly	Ala	Val	
	465					470					475					480	
	CAG	AAC	RAG	GTT	RCC	CTY	ACA	CAC	CCY	ATA	ACC	AAG	TWC	ATC	ATG	RCA	1488
45	Gln	Asn	Xak	Val	Xal	Leu	Thr	His	Pro	Ile	Thr	Lys	Xam	Ile	Met	Xan	
15					485					490					495		
	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	TGG	GTG	CTG	GTA	1536
	Сув	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	Val	
				500					505					510			
20	GGC	GGG	GTC	CTC	GCG	GCT	CTG	GCC	GCG	TAC	TGC	CTG	ACA	ACG	GGC	AGC	1584
	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Сув	Leu	Thr	Thr	Gly	Ser	
			515					520					525				
	GTG	GTC	ATT	GTG	GGC	AGG	ATC	ATC	TTG	TCC	GGG	AGG	CCG	GCC	GTT	ATT	1632
25	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	Val	Ile	
		530					535					540					
	CCC	GAC	AGG	GAA	GTT	CTC	TAC	CAA	GAG	TTC	GAT	GAA	ATG	GAA	GAG	TGC	1680
	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	Cys	
30	545					550					555				-	560	
					CCT												1728
	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu	Gln	
					565					570					575		
35					GCG												1776
	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	Gln	Ala	
				580					585					590			
	GAG	GCT	GCT	GCT	CCC	GTG	GTG	GAG	TCC	AAG	TGG	CGA	GCC	CTT	GAG	ACC	1824
40	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	Glu	Thr	
			595					600					605				
	TTC	TGG	GCG	AAG	CAC	atg	TGG	AAT	TTC	ATC	AGC	GGG	ATA	CAG	TAC	TTA	1872
	Phe	$\mathtt{Trp}$	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	Tyr	Leu	
45		610					615					620					
	GCA	GGC	TTG	TCC	ACT	CTG	CCT	GGA	AAC	CCC	GCA	ATA	GCA	TCA	CTG	ATG	1920

	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	Ala	Ser	Leu	Met	
	625					630					635					640	
	GCA	TTC	ACA	GCC	TCT	ATC	ACC	AGC	CCG	CTC	ACC	ACC	CAA	TAT	ACC	CTC	1968
5	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	Gln	Tyr	Thr	Leu	
					645					650					655		
		TTT															2016
	Leu	Phe	Asn	Ile	Leu	Gly	Gly	Trp	Val	Ala	Ala	Gln	Leu	Ala	Pro	Pro	
10				660					665					670			
		GCC															2064
	Ser	Ala		Ser	Ala	Phe	Val	Gly	Ala	Gly	Ile	Ala	Gly	Ala	Ala	Val	
			675					680					685				
15		AGC															2112
	Gly	Ser	Ile	Gly	Leu	Gly		Val	Leu	Val	Asp	Ile	Leu	Ala	Gly	Tyr	
		690					695					700					
		GCA															2160
20		Ala	Gly	Val	Ala		Ala	Leu	Val	Ala	Phe	Lys	Val	Met	Ser	Gly	
	705					710					715			•		720	
		ATG															2208
	Asp	Met	Pro	Ser		Glu	Asp	Leu	Val	Asn	Leu	Leu	Pro	Ala	Ile	Leu	
25					725	_				730					735		
		CCT															2256
	ser	Pro	GIY		Leu	Val	Val	Gly		Val	Cys	Ala	Ala		Leu	Arg	
	ccc	Cam	CMC	740	003	000			745		_			750			
30		CAT															2304
	ALG	His	755	GIĀ	PIO	GIY	GIU		Ala	Val	GIn	Trp		Asn	Arg	Leu	
	v :	C or	_		R:	<b>A</b> 01	- C	760	v		6		765		_	_	
		Gor			W :					: A						or T	
35		: Le	-			A OI		h .		: A or A		. or				or T	_
		: Se								or A	_					ror	
		: Th							_	or I	-					e or	
		: Ty								or L						o or	
40	-	: Ty								or A	_	•		vaT	· Tn	r or	ATG
		- 1							-111	OT W							

SEQ ID NO:43

SEQUENCE LENGTH: 3564 base pairs

45 SEQUENCE TYPE: nucleic acid

55

STRANDEDNESS: double

Br. J. W. J.

	TOPOLO	GY:	line	ar													
_	ANTI-S	ENSE	: No														
5	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa	titi	s C	viru	8										
	IMMEDI.	ATE :	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	MX2	5ท15														
10																	
	TGT	GCC	TGG	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAA	GCT	GAG	GCC	GCC	48
	Сув	Ala	Trp	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	
	1				5					10					15		
15	TTG	GAG	AAC	CTG	GTG	GTC	CTC	AAT	GCA	GCA	TCC	ATG	GCS	GGA	GCG	CAT	96
	Leu	Glu	Asn	Leu	Val	Val	Leu	Asn	Ala	Ala	Ser	Met	Ala	Gly	Ala	His	
				20					25					30			
00						CTT											144
20	Gly	Ile	Leu	Ser	Phe	Leu	Val	Phe	Phe	Cys	Ala	Ala	Trp	Tyr	Ile	Lys	
	σ.		35					40	•				45				
						GGG											192
25	Gly		Leu	Val	Pro	Gly		Xaa	Tyr	Ala	Xab	Tyr	Gly	Val	Trp	Pro	
25		50					55					60					
						MTG											240
		Leu	Leu	Leu	Leu	Xac	Ala	Leu	Pro	Xad		Ala	Tyr	Ala	Met		
30	65	010	3.00	aam		70					75					80	
30						TCG											288
	ALY	Ade	met	AIG	A1A 85	Ser	Cys	GIY	GIY		Val	Phe	Val	Gly		Val	
	ርሞር	VIVC	NCC.	mmc		CCA	m> 0	m> a		90					95		
35						CCA											336
••	Dou	Dea	1111	100	Ser	Pro	TYL	TYL	105	vaı	Pne	ren	ATA		Leu	IIe	
	TGG	TGG	מיזיים		ጥልጥ	CTC	እጥሮ	»CC		ccc	CAC	ccc		110	~~~	cma	204
						Leu											384
40			115		-,-	Lou		120	ALG	VIC	GIU	ATG	125	neu	GIII	vai	
-	TGG	ATY		CCY	CTY	AAC	GTY		GGR	GGC	CGC	GAV		ልጥሮ	አጥሮ	CWA	432
						Asn											432
	•	130					135	9	1	1	9	140		**6	116	Deu	
45	СТС	ACR	TGT	GCG	GTC	CAY		GAG	CTR	АТУ	ТФТ		ATC	ACC	AAR	CTV	480
-						His											400
	-		. 4 .								- 11G	יוטני	-1G	****	ny a	Ten	

55

	145					150					155					160	
	YTG	CTC	GCC	ATA	CTC	GGT	CCG	CTC	ATG	GTR	CTC	CAG	GCT	GSC	MTA	ACY	528
-	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Xag	Xah	Thr	
5					165					170					175		
	MRA	RTG	CCG	TAC	TTY	GTR	CGY	GCT	CAA	GGG	CTC	ATY	CGT	RYG	TGC	ATG	576
	Xai	Xaj	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Xak	Сув	Met	
10				180					185					190			
10	TTR	GTG	CGG	AAA	GYC	GCY	GGR	GGT	CAT	TAT	GTY	CAR	ATG	GCY	YTY	RTG	624
	Leu	Val	Arg	Lys	Xal	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Xam	Xan	
			195					200					205				
15	AAG	CTG	GCY	GCR	CTG	ACA	GGT	ACG	TAC	RTT	TAT	GWC	CAT	CTT	RCY	CCA	672
75	Lys	Leu	Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Xao	Tyr	Xap	His	Leu	Xaq	Pro	
		210					215					220					
	CTG	CAG	SAY	TGG	GCC	CAY	GCG	GGC	CTA	CGR	GAC	CTT	GCG	GTR	GCR	GTW	720
20	Leu	Gln	Xar	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	
20	225					230					235					240	
	GAG	ccc	GTT	GYC	TTC	TCT	GAY	ATG	GAG	ACY	AAG	ATC	ATC	ACS	TGG	GGG	768
	Glu	Pro	Val	Xas	Phe	Ser	Asp	Met	Glu	Thr	Lys	Ile	Ile	Thr	Trp	Gly	-
25					245					250			-		255		
20	GCA	GAS	ACB	GCG	GCG	TGT	GGG	GAC	ATC	ATY	TYG	GGY	CTA	ССН	GTY	TCC	816
	Ala	Xat	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Xau	Gly	Leu	Pro	Val	Ser	
				260					265					270			
30						GAG											864
<b>50</b>	Ala	Arg	_	Gly	Xav	Glu	Xaw		Xax	Gly	Pro	Ala	_	Ser	Phe	Asp	
			275				_	280					285				
						CTC									_		912
35	GTĀ		GLY	Trp	Arg	Leu		Ala	Pro	He	Thr		TYY	Ser	GIn	Gin	
00	300	290	~~~				295					300		~~~	222	<b></b>	0.50
						GGT											960
		Arg	GLY	Leu	Leu	Gly	Cys	TTE	me	Thr		Leu	Thr	GLY	Arg		
40	305		~~~			310					315					320	1000
**						GGG											1008
	гÀв	Asn	GIN	vaı		Gly	GIU	vai	GIN		vaı	ser	Tnr	AIA		GIN	
	m/°m	mm.c	ama	CCC	325	m/s**	DMC	330	ccc	330	mcc.	maa	N COM	Cittor	335	CAC	1056
45						TGY											1056
70	oel	rne	теп		THE	Суѕ	xay	ABN	_	val	Cys	rrp	THE		rne	ura	
				340					345					350			

	CCV	ccc	ccc	ሞርር	AAG	NCC.	mm x	ccc	ccc	CCA		cco		NMC	300	C) )	1104
																	1104
	GLY	Ald	355	Ser	Lys	TIIT	nea	360	GIY	PLO	гЛя	GIA	365	TTE	THE	GIII	
5	рпа	ሞልሮ		<b>ከል</b> ወ	GTR	СУТ	ርልር		ሮሞሮ	CTC	CCV	ሞርር		ece	ccc	ccc	1152
					Val												1152
		370			vul	nop	375	nop	neu	VOI	GIY	380	Ser	YIG	PLO	PLO	
	SGG		ССТ	TCC	TTG	ACA		ጥርሮ	ACC	ጥርር	ממר		ምሮር	GAC	Calous	ጥአጥ	1200
10					Leu												1200
	385	****	9	001		390	110	Cyb	1111	Cyb	395	Ser	Ser	nsp	Deu	400	
		GTC	ACG	AGR	CAT		САТ	СПС	አ ጥጥ	CCG		CAC	ccc	ccc	ccc		1248
					His												1240
15				9	405	••••	шр		110	410	vul	111.0	n.y	ary	415	vob	
	AGC	AGG	GGG	AGC	CTS	CTS	TCS	CCC	GGG		ATC	TCY	TÁC	YTE		CCV	1296
					Leu												1270
		-		420					425			552	-1-	430		0-1	
20	TCC	TCG	GGT		CCG	CTG	CYT	TGC		TCG	GGC	CRT	GTT		GGC	ATC	1344
					Pro												
			435	•				440			1		445		1		
	TTC	CGG	GCT	GCY	GTG	TGC	ACC	CGG	GGG	GTT	GCG	AAG		GTR	GAC	TTT	1392
25					Val												
		450				-	455	_	-			460					
	GTG	CCC	GTT	GAG	TCT	ATG	GAA	ACC	ACY	ATG	CGG	TCT	CCG	GTC	TTC	RCG	1440
					Ser												
30	465					470					475					480	
	GAT	AAC	TCA	ACC	CCC	CCG	GCC	GTA	CCG	CAG	WCA	TTC	CAA	GTG	GCC	CAC	1488
	Asp	Asn	Ser	Thr	Pro	Pro	Ala	Val	Pro	Gln	Xbd	Phe	Gln	Val	Ala	His	
					485					490					495		
35	CTA	CAC	GCT	CCC	ACT	GGC	AGC	GGC	AAA	AGC	ACC	ARG	GTG	CCG	GCT	GCG	1536
	Leu	His	Ala	Pro	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Xbe	Val	Pro	Ala	Ala	
				500					505					510			
	TAT	GCG	GCC	CAA	GGG	TAC	AAG	GTA	CTC	GTC	CTG	AAC	CCG	TCC	GTT	GCT	1584
40	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	
			515					520					525				
	GCC	ACT	TTG	GGC	TTT	GGG	GCG	TAY	ATG	TCC	AAG	GCA	CAT	GGT	GTT	GAC	1632
	Ala	Thr	Leu	Gly	Phe	${\tt Gly}$	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Val	Asp	
45		530					535					540					
	CCT	AAC	ATC	AGA	ACT	GGG	GTG	AGG	ACC	ATC	ACC	ACG	GGC	GCT	CCC	RTC	1680

	Pro	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro	Xbf	
	545					550					555		_	•		560	
_	ACG	TAC	TCC	ACC	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	1728
5	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Сув	Ser	Gly	
					565					570			_	-	575	-	
	GGT	GCC	TAT	GAC	ATC	ATA	ATA	TGT	GAT	GAG	TGT	CAT	TCA	ACT	GAC	TCG	1776
	Gly	Ala	Tyr	Asp	Ile	Ile	Ile	Cys	Авр	Glu	Cys	His	Ser	Thr	Asp	Ser	
10				580					585					590			
	ACT	TCC	ATC	TTG	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG	GAG	ACG	GCT	1824
	Thr	Ser	Ile	Leu	${\tt Gly}$	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	
			595					600					605				
15	GGA	GCG	CGC	CTT	GTC	GTG	CTC	GCC	ACC	GCT	ACG	CCT	CCG	GGA	TCG	GTC	1872
	Gly	Ala	Arg	Leu	Val	Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	
		610					615					620					
	ACC	GTG	CCG	CAT	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA	1920
20		Val	Pro	His	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Asn	Thr	Gly	
	625					630					635					640	
															AAG		1968
0.5	Glu	Ile	Pro	Phe	Tyr	Gly	Lys	Ala	Ile	Pro	Leu	Glu	Ala	Ile	Lys	Gly	
25					645					650					655		
															GAG		2016
	Gly	Arg	His		Xbg	Phe	Сув	His		Lys	Lys	Lys	Сув	Asp	Glu	Leu	
00				660					665					670			
30															TAC		2064
	А1а	Ala		Leu	Ser	Ala	Leu		Val	Asn	Ala	Val		Tyr	Tyr	Arg	
		~	675					680					685	_			
35															GTC		2112
33	GIĀ	690	Asp	vai	ser	XDN		Pro	Thr	ser	GIĀ	_	Val	Val	Val	Val	
	CCA		CAC	CCM	COL	አመሮ	695	cca	mam.	3.00	0011	700	mmrr	~~~	<b>~~</b>	ama	01.60
															TCA		2160
40	705	1111	мэр	Ата	neu	710	THE	GIY	TYL	Thr		Авр	Pne	Asp	Ser		
40		CAC	mcc	770	አሮአ		CMC	200	CAA	303	715 CTC	C3M	mma	300	TTG	720	2200
															Leu		2208
		-10 P	C. 7 3	13011	725	cya	401	*111T	GIII	730	491	waħ	FIIE	ser		wab	
45	ССТ	АСТ	ባጥር	ACC		GAG	<b>A</b> CG	እ <b>ሮ</b> ር	<b>ACC</b>		CCC	C D D	ርስም	ccc	735 GTG	mcc.	2256
10															Val		2256
			1116	TILL	116	GIU	TIIT	1117	TIIL	val	PLU	GTII	vaħ	wig	AGT	ser	

				740					745					750			
	CGC	TCG	CAG	CGG	CGA	GGC	AGG	ACT	GGT	AGG	GGC	AGR	GGG	GGC	ATA	TAC	2304
-	Arg	Ser	Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	
5			755					760					765				
	AGG	TTT	GTA	ACT	CCA	GGG	GAA	CGG	ccc	TCA	GGC	ATG	TTC	GAT	TCT	TCG	2352
	Arg	Phe	Val	Thr	Pro	Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	
	•	770					775					780					
10	GTC	CTG	TGT	GAA	TGT	TAT	GAC	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACG	2400
	Val	Leu	Cys	Glu	Cys	Tyr	Asp	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	
	785					790					795					800	
	YCC	GCC	GAG	ACC	TCG	GTT	AGG	TTG	CGG	GCT	TAC	CTA	AAY	ACA	CCT	GGG	2448
15	Xbi	Ala	Glu	Thr	Ser	Val	Arg	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	
					805					810					815		
	CTG	CCC	GTC	TGC	CAG	GAC	CAT	CTG	GAG	TTC	TGG	GAG	AGC	GTC	TTC	ACC	2496
	Leu	Pro	Val	Cys	Gln	Asp	His	Leu	Glu	Phe	Trp	Glu	Ser	Val	Phe	Thr	
20				820					825					830			
	GGC	CTC	ACC	CAC	ATA	GAT	GCC	CAC	TTC	TTG	TCC	CAG	ACY	AAA	CAG	GCA	2544
	Gly	Leu	Thr	His	Ile	Asp	Ala	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	
			835					840					845				
25	GGA	GAC	AAC	TTC	ccc	TAC	CTG	GTA	GCA	TAC	CAG	GCT	ACA	GTG	TGC	GCC	2592
	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	Сув	Ala	
		850		-			855					860					
	AGG	GCC	AAG	GCT	CCA	CCT	CCA	TCG	TGG	GAT	CAR	ATG	TGG	AAG	TGT	CTC	2640
30	Arg	Ala	Lys	Ala	Pro	Pro	Pro	Ser	Trp	Asp	Gln	Met	Trp	Lys	Cys	Leu	
	865					870					875					880	
	ATA	CGG	CTG	AAG	CCT	ACG	CTA	CAC	GGG	CCA	ACG	CCC	CTG	TTG	YAT	AGG	2688
	Ile	Arg	Leu	Lys	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Xbj	Arg	
35					885					890					895		
						AAC											2736
	Leu	Gly	Ala	Val	Gln	Asn	Xbk	Val	Xbl	Leu	Thr	His	Pro	Ile	Thr	Lys	
40				900					905					910			
	TWC	ATC	ATG	RCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	2784
	Xbm	Ile	Met	Xbn	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	
			915					920					925				
45						GGG											2832
	Trp		Leu	Val	Gly	Gly		Leu	Ala	Ala	Leu		Ala	Tyr	Сув	Leu	
		930					935					940			•		

	ACA	ACG	GGC	AGC	GTG	GTC	ATT	GTG	GGC	AGG	ATC	ATC	TTG	TCC	GGG	AGG	2880
				s r													
	945					950					955				-	960	
5	CCG	GCC	GTT	ATT	CCC	GAC	AGG	GAA	GTT	CTC	TAC	CAA	GAG	TTC	GAT	GAA	2928
				Ile													
					965					970					975		
	ATG	GAA	GAG	TGC	GCC	TCG	CAC	CTC	ССТ	TAC	ATC	GAA	CAA	GGA	ATG	CAG	2976
10	Met	Glu	Glu	Cys	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Glņ	Gly	Met	Gln	
				980					985					990			
	CTC	GCC	GAG	CAA	TTC	AAG	CAG	AAG	GCG	CTC	GGT	TTG	CTG	CAA	ACA	GCC	3024
	Leu	Ala	Glu	Gln	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	
15			995					1000					1005				
				GCG													3072
	Thr	Lys	Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	
		1010					1015					L020					
20				ACC													3120
		Leu	Glu	Thr	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	
	1025					1030					1035					1040	
				TTA													3168
25	Ile	Gln	Tyr	Leu		Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	
					1045					L050					.055		
				ATG													3216
		Ser		Met	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	
-30		 		1060					L065					1070	-		•
				CTC													3264
	GIN			Leu	ren	Phe			Leu	Gly	Gly			Ala	Ala	Gln	
	CITIC		075	000	3.cm			080					.085				
35				CCC													3312
		L090	PIO	Pro	ser			Ser	Ala	Phe			Ala	Gly	Ile	Ala	
			C C m	C/Pm	000		.095					.100					
				GTT													3360
40		Ald	Ala	Val			He	GLY	Leu			Val	Leu	Val	Asp	Ile	
	1105	ccc	ccm	m» m		110					115					.120	
				TAT													3408
	neg	wid	GTĀ	Tyr		ATS	стλ	vaı			Ala	Leu	Val			Lys	
<b>4</b> 5	ርመር	አጥር	NCC.		.125	N IDICO	000	maa		130					135		
	GIC	TIG	AGC	GGT	GAC	AIG	CCC	TCC	ACC	GAG	GAC	CTG	GTC	AAC	TTA	CTC	3456

	Val Met Ser Gly Asp Met Pro Ser Thr Glu Asp Leu Val Asn Leu Leu	
	1140 1145 1150	
5	CCC GCC ATC CTC TCT CCT GGT GCC CTG GTC GTC GGG GTC GTG TGC GCA 3504	ı
	Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala	
	1155 1160 1165	
	GCA ATA CTG CGT CGG CAT GTG GGC CCA GGG GAG GGG GCT GTG CAG TGG 3552	2
10	Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp	
	1170 1175 1180	
	ATG AAC CGG CTG 3564	Ī
	Met Asn Arg Leu	
15	1185	
	Y:CorT R:AorG M:AorC K:GorT	
	S:GorC W:AorT H:AorCorT B:GorTorC	;
	Xaa : Ala or Thr Xab : Phe or Leu Xac : Met or Leu	
20	Xad: Ala or Pro Xae: Glu or Asp Xaf: Lys or Arg	
	Xag : Gly or Ala Xah : Leu or Ile Xai : Gln or Arg	
	Xaj: Met or Val Xak: Met or Ala Xal: Ala or Val	
	Xam: Leu or Phe Xan: Met or Val Xao: Val or Ile Xap: Asp or Val Xag: Thr or Ala Xar: Asp or His	
25	Vac a Name of the control of the con	
	Xas: Ala or Val Xat: Asp or Glu Xau: Leu or Ser Xav: Asn or Arg or Lys Xaw: Ile or Leu Xax: Leu or Phe	
	Xay: Ile or Val Xaz: Gly or Arq Xba: Leu or Pro	
	Xbb: His or Arg Xbc: Thr or Ala Xbd: Ser or Thr	
30	Xbe: Lys or Arg Xbf: Ile or Val Xbg: Thr or Ile	
	Xbh : Val or Ile Xbi : Pro or Ser Xbj : Tyr or His	
	Xbk : Glu or Lys Xbl : Thr or Ala Xbm : Tyr or Phe	
	Xbn : Thr or Ala	
35		
	SEQ ID NO:44	
	SEQUENCE LENGTH: 849 base pairs	
	SEQUENCE TYPE: nucleic acid	
40	STRANDEDNESS: double	
	TOPOLOGY: linear	
	ANTI-SENSE: No	
	ORIGINAL SOURCE	
45	ORGANISM: Hepatitis C virus	

50

IMMEDIATE EXPERIMENTAL SOURCE

### CLONE: MX25-1

5	TGI	GCC	TGG	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAA	GCI	GAG	GCC	GCC	48
Ū	Cys	Ala	Trp	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	
	1				5					10	i				15	•	
	TTG	GAG	AAC	CTG	GTG	GTC	CTC	AAT	GCA	GCA	TCC	ATG	GCG	GGA	GCG	CAT	96
10	Leu	Glu	Asn	Leu	Val	Val	Leu	Asn	Ala	Ala	Ser	Met	Ala	Gly	Ala	His	
				20					25					30			
	GGC	ATC	CTC	TCT	TTC	CTT	GTG	TTC	TTC	TGT	GCC	GCC	TGG	TAC	ATC	AAA	144
	Gly	Ile	Leu	Ser	Phe	Leu	Val	Phe	Phe	Сув	Ala	Ala	Trp	Tyr	Ile	Lys	
15			35					40	•				45				
,,	GGC	AGG	CTG	GTC	CCT	GGG	GCG	GCA	TAC	GCT	TTC	TAT	GGC	GTA	TGG	CCG	192
	Gly	Arg	Leu	Val	Pro	Gly	Ala	Ala	Tyr	Ala	Phe	Tyr	Gly	Val	Trp	Pro	
		50					55					60					
20	CTG	CTC	CTG	CTC	TTG	ATG	GCG	CTA	CCC	GCA	CGG	GCG	TAC	GCC	ATG	GAC	240
20	Leu	Leu	Leu	Leu	Leu	Met	Ala	Leu	Pro	Ala	Arg	Ala	Tyr	Ala	Met	Asp	
	65					70					75					80	
					GCA												288
05	Arg	Glu	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	Leu	Val	
25					85					90					95		
					TCA												336
	Leu	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Val	Phe	Leu	Ala	Lys	Leu	Ile	
				100					105					110			
30	TGG	TGG	TTG	CAA	TAT	CTC	ATC	ACC	AGG	GCC	GAG	GCG	CAC	TTG	CAA	GTG	384
	Trp	Trp	Leu	Gln	Tyr	Leu	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	Gln	Val	
			115					120					125	•			
•	TGG	ATC	CCC	CCC	CTC	AAC	GTT	CGG	GGG	GGC	CGC	GAT	GCC	ATC	ATC	CTT	432
35	Trp		Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Ile	Ile	Leu	
		130					135					140					
	CTC	ACA	TGT	GCG	GTC	CAC	CCG	GAG	CTG	ATC	TTT	GAC	ATC	ACC	AAG	CTC	480
		Thr	Cys	Ala	Val	His	Pro	Glu	Leu	Ile	Phe	Asp	Ile	Thr	Lys	Leu	
40	145					150					155					160	
					CTC												528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Gly	Leu	Thr	
					165					170					175		
45	CAA	ATG	CCG	TAC	TTT	GTG	CGT	GCT	CAA	GGG	CTC	ATT	CGT	ATG	TGC	ATG	576
	Gln	Met	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Met	Cys	Met	

50

				180					185					190			
	TTG	GTG	CGG	AAA	GCC	GCT	GGG	GGT	CAT	TAT	GTC	CAG	ATG	GCT	CTC	ATG	624
5		Val															
			195					200					205				
	AAG	CTG	GCT	GCA	CTG	ACA	GGT	ACG	TAC	GTT	TAT	GAC	CAT	CTT	ACT	CCA	672
	Lys	Leu	Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Asp	His	Leu	Thr	Pro	
10	•	210					215					220					
-		CAG															720
		Gln	Asp	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	
	225					230					235					240	
15		CCC															768
	Glu	Pro	Val	Ala	Phe	Ser	Asp	Met	Glu	Thr	Lys	Ile	Ile	Thr	Trp	Gly	
					245					250					255		
		GAC															816
20	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro	Val	Ser	
				260					265					270			
		CGG															849
	Ala	Arg		Gly	Asn	Glu	Ile	Leu	Leu	Gly	Pro						
25			275					280				-					
			_														
	SEQ ID																
	SEQUEN							5									
30	SEQUEN					acio	i										
	STRANDI				.e												
	TOPOLO			ar													
	ANTI-SI			_													
35	ORIGINA			_													
	ORGANIS																
	IMMEDIA CLONE:			(IME)	TAL	SOUL	CE										
	CHOME.	PIAZ	-2														
40	ጥርጥ	GCC	ጥርር	መጥር	ጥርር	አጥር	አመሮ	CMC	CMC	».	cco	0	aam	<b>a</b> .a	000		4.0
		Ala															48
	1			Deu	5	nec	Mec	neu	neu	10	WIG	GIII	ATG	GIU		Ala	
	_	GAG	AAC	CTC	_	ርሞር	ርሞር	መልል	CCA		mee	አሙጣ	ccc	CCX	15	O a m	0.0
45		Glu															96
				20		· u i	ac u	HOII	25	WTG	PET	Met	wrq	_	WIG	urs	
				20					45					30			

	GGC	ATC	CTC	TCT	TTC	CTT	GTG	TTC	TTC	TGT	GCC	GCC	TGG	TAC	ATC	AAA	144
					Phe												
_			35					40		_			45	-		-	
5	GGC	AGG	CTG	GTC	CCT	GGG	GCG	ACA	TAC	GCT	CTC	TAT	GGC	GTA	TGG	CCG	192
	Gly	Arg	Leu	Val	Pro	Gly	Ala	Thr	Tyr	Ala	Leu	Tyr	Gly	Val	Trp	Pro	
		50					55					60					
10					TTG												240
70	Leu	Leu	Leu	Leu	Leu	Met	Ala	Leu	Pro	Pro	Arg	Ala	Tyr	Ala	Met	Asp	
	65					70					75					80	
					GCA												288
15	Arg	Asp	Met	Ala	Ala	Ser	Cys	Gly	Gly		Val	Phe	Val	Gly	Leu	Val	
	~~~				85					90					95		
					TCA												336
	ren	ren	Thr		Ser	Pro	Туг	Tyr		Val	Phe	Leu	Ala	-	Leu	Ile	
20	mcc.	mcc	mma	100	mam	CMC	» m.c	200	105	000	a.a	000	a.a	110		ama	204
					TAT												384
	TTP	ııp	115	GIII	Tyr	neu	116	120	ALG	AIG	GIU	AIA	125	Leu	GIN	val	
	TGG	ΑͲͲ		ССТ	CTC	אאר	ርጥር		CCA	GGC	ccc	CAC		አጥሮ	እመሮ	CTC	432
25					Leu												732
	•	130					135		1	- 1	9	140				200	
	CTC	ACG	TGT	GCG	GTC	CAT	CCA	GAG	CTA	ATT	TTT	GAC	ATC	ACC	AAA	CTT	480
					Val												
30	145					150					155	·)			_	160	
	CTG	CTC	GCC	ATA	CTC	GGT	CCG	CTC	ATG	GTG	CTC	CAG	GCT	GCC	ATA	ACT	528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Ala	Ile	Thr	
					165					170					175		
35	AGA	GTG	CCG	TAC	TTC	GTA	CGC	GCT	CAA	GGG	CTC	ATC	CGT	GCG	TGC	ATG	57 6
	Arg	Val	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Ala	Cys	Met	
				180					185					190			
					GCC												624
40	Leu	Val		Lys	Ala	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Phe	Val	
			195					200					205				
					CTG												672
	гăа		Ala	Ala	Leu	Thr		Thr	Tyr	Ile	Tyr		His	Leu	Ala	Pro	
45	CMC	210	G3.E	mac	000		215					220					
	CIG	CAG	CAT	TGG	GCC	CAT	GCG	GGC	CTA	CGG	GAC	CTT	GCG	GTG	GCG	GTA	720

			His	Trp	Ala		Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	
	225					230					235					240	
5			GTT														768
J	GLu	Pro	Val	Val		Ser	Asp	Met	Glu	Thr	Lys	Ile	Ile	Thr	Trp	Gly	
					245					250					255		
			ACC														816
40	Ala	Asp	Thr		Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro	Val	Ser	
10				260					265					270			
			AGG														849
	Ala	Arg	Arg	Gly	Asn	Glu	Ile	Leu	Leu	Gly	Pro						
			275					280									
15																	
	SEQ ID																
	SEQUEN					_		Б								•	
	SEQUEN					acio	i										
20	STRAND	EDNE	SS: c	loub.	le												
	TOPOLO			ar													
	ANTI-S																
	ORIGIN		_														
25	ORGANI		_														
	IMMEDIA			RIME	LATV	SOU	RCE										
	CLONE:	MX2	5-3														
30			TGG														48
		Ala	Trp	Leu		Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	
	1				5					10					15		
			AAC														96
35	ren	GIU	Asn		Val	Val	Leu	Asn		Ala	Ser	Met	Ala		Ala	His	
	999			20					25					30			
			CTC														144
	GIĀ	iie	Leu	Ser	Phe	Leu	Val		Phe	Cys	Ala	Ala	Trp	Tyr	Ile	Lys	
40	~~~		35					40				•	45				
			CTG														192
	GTĀ		Leu	Val	Pro	Gly		Ala	Tyr	Ala	Phe	Tyr	Gly	Val	Trp	Pro	
	200	50		-			55					60					
45			CTG														240
	тел	тел	Leu	Leu	Leu	Leu	Ala	Leu	Pro	Ala	Arg	Ala	Tyr	Ala	Met	Asp	•

	65					70					75					80	
	CGG	GAG	ATG	GCT	GCA	TCG	TGC	GGA	GGC	GCG	_	ጥጥጥ	GTA	ርርጥ	CTG	GTA	288
5					Ala												200
J					85		-	-		90				1	95	,	
	CTC	CTG	ACC	TTG	TCA	CCA	TAC	TAC	AAA		TTC	СТС	GCT	AAG		מיזים	336
					Ser												330
40	•			100			-	•	105					110	200		
10	TGG	TGG	TTG	CAA	TAT	CTC	ATC	ACC	AGG	GCC	GAG	GCG	CAC	_	CAA	GTG	384
					Tyr												501
			115					120	_				125				
	TGG	ATC	CCC	CCC	CTT	AAC	GTT	CGG	GGG	GGC	CGC	GAT	GCC	ATC	ATC	CTT	432
15					Leu												
		130					135			_		140					
	CTC	ACA	TGT	GCG	GTC	CAC	CCG	GAG	CTG	ATC	TTT	GAC	ATC	ACC	AAG	CTC	480
00					Val												
20	145					150					155				_	160	
	TTG	CTC	GCC	ATA	CTC	GGT	CCG	CTC	ATG	GTA	CTC	CAG	GCT	GGC	CTA	ACC	528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Gly	Leu	Thr	
					165					170					175		
25	CAA	ATG	CCG	TAC	TTT	GTG	CGT	GCT	CAA	GGG	CTC	ATT	CGT	ATG	TGC	ATG	576
	Gln	Met	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Met	Сув	Met	
				180					185					190			•
20					GTC												624
30	Leu	Val	Arg	Lys	Val	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Leu	Met	
			195					200					205				
					CTG												672
35	Lys		Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Val	His	Leu	Thr	Pro	
33		210					215					220					
					GCC												720
	· Leu	Gln	Asp	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	
40	225					230					235					240	
40					TTC												768
	GIu	Pro	Val	Val	Phe	Ser	Asp	Met	Glu		Lys	Ile	Ile	Thr	Trp	Gly	
		a			245					250					255		
AE.					GCG												816
45	ATG	Asp	rnr		Ala	Cys	Gly	Asp		Ile	Leu	Gly	Leu		Val	Ser	
				260					265				•	270			

	GCC	CGG	AGG	GGC	AAC	GAG	ATA	CTC	CTC	GGA	CCG						849
	Ala	Arg	Arg	Gly	Asn	Glu	Ile	Leu	Leu	Gly	Pro						
5			275					280									
3																	
	SEQ ID	NO:	47														
	SEQUEN	ICE L	ENGT	H: 5	24 b	ase 1	pair	в						•			
10	SEQUEN	ICE T	YPE:	nuc	leic	acio	i										
10	STRAND	EDNE	SS: (doub	le												
	TOPOLO	GY:	line	ar													
	ANTI-S	ENSE	: No													,	
15	ORIGIN	AL S	OURC	E						-							
15	ORGANI	SM:	Hepa	titi	в С 1	virus	3										
	IMMEDI	ATE	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	026	-1														
20																	
20			TGG														48
			Trp	Gly		Glu	Thr	Ala	Ala		Gly	Asp	Ile	Ile	Ser	Gly	
	1				5					10					15		
25			GTT														96
20	Leu	Pro	Val		Ala	Arg	Arg	Gly		Glu	Leu	Leu	Leu		Pro	Ala	
	G N m		mmm	20		~~~			25					30			
			TTT														144
30	Asp	ser	Phe	Asp	GLY	GIN	GTĀ		Arg	Leu	Leu	ATa		Ile	Thr	Ala	
00	ma c	mcc	35	CAC	300	000	000	40	~mm	~~~			45				
			CAG														192
	+y-	50	Gln	GIII	THE	ALG	55	nen	Leu	GIY	Cys	90	TTE	THE	Ser	ren	
35	ACG		CGG	ርልጥ	DAG	אממ		CTYC	GNG	ccc	GAG		CAA	CITC.	CTIC	mem	240
			Arg														240
	65		1119	2100	ay o	70	GIN	141	Giu	GLY	75	441	GIII	Val	Vai	80	
			ACA	CAA	ጥርጥ		СТС	GCG	ልሮሮ	ጥርጥ	_	ממ	CCC	CTC	TICC		288
40			Thr														200
				U	85		204	****	1111	90	741	2111	323	VUI	95		
	ACT	GTT	TTC	CAC		GCC	GGC	TCG	AAG		מיויים	GCC	GGC	CCA		GGC	336
			Phe														330
45				100	1		1		105				0-1	110	-,,	O.J	
	CCA	ATC	ACC	CAA	ATG	TAC	ACC	AAT		GAT	CAG	GAC	CTC		GGC	TGG	384
					-	_											
50																	

	Pro	Ile	Thr	Gln	Met	Tyr	Thr	Asn 120	Val	Asp	Gln	Asp	Leu 125	Val	Gly	Trp	
_	TCG	GCG	ccc	ccc	CGG	GCG	CGT		TTG	ACA	ССТ	ሞፍሮ		ጥርር	ccc	NGC	432
5		Ala															432
		130			•		135					140		Cyb	GLY	Ser	
	TCG	GAC	CTT	TAT	TTG	GTC		AGG	CAT	GCT	САТ		ጀ ባንጥ	CCG	GTG	CAC	480
		Asp															400
10	145			-		150		9			155	, u	110	110	AGI	160	
		CGG	GGC	GAC	AGC		GGG	AGC	ርሞር	רידיר		CCC	ccc	ccc	λm	100	524
		Arg													N.		324
	,	•			165	9	1	002		170	001	110	GLY	FIO			
15										1,0							
	SEQ ID	NO:	18														
	SEQUEN	CE LI	ENGTI	H: 5	14 ba	ase 1	pair	6				•					
	SEQUEN							_									
20	STRAND																
	TOPOLO	GY:]	linea	ar													
	ANTI-S	ENSE:	: No														
	ORIGIN	AL S	OURCE	3									•				
25	ORGANI	SM: I	lepat	itis	3 C 1	rirus	3										
	IMMEDI																
	CLONE:	026-	-2														
30	ATC	ACG	TGG	GGG	GCA	GAG	ACG	GCG	GCG	TGT	GĞG	GAC	ATC	ATC	TCG	GGT	48
		Thr															
	1				5					10	_	-			15	4	
	CTA	CCC	GTT	TCC	GCC	CGA	AGG	GGG	AGG	GAG	CTG	CTT	TTG	GGA	CCG	GCC	96
35		Pro															
				20					25					30			
	GAT	AGT	TTT	GAC	GGG	CAG	GGG	TGG	CGA	CTC	CTT	GCG	ССТ	ATC	ACG	GCC	144
	Asp	Ser	Phe	qaA	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	
40			35					40					45				
	TAC	TCC	CAG	CAG	ACG	CGG	GGC	CTG	CTT	GGT	TGC	ATC	ATC	ACC	AGC	CTT	192
	Tyr	Ser	Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	
		50					55					60					
45	ACG	GGC	CGG	GAT	AAG	AAC	CAG	GTC	GAG	GGG	GAG	GTT	CAA	GTG	GTC	TCT	240
		Gly															
										_							
50																	
50																	

	65					70					75					80	
	ACC	GCA	ACA	CAA	TCT	TTC	CTG	GCG	ACC	TGC	ATC	AAC	GGC	GTT	TGC	TGG	288
5					Ser												
J					85					90			,		95	•	
	ACT	GTT	TTC	CAC	GGC	GCC	GGC	TCG	AAG	ACC	TTA	GCC	GGC	CCA	AAA	GGC	336
	Thr	Val	Phe	His	Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	
10				100		•			105					110			
					ATG												384
	Pro	Ile		Gln	Met	Tyr	Thr	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	
			115					120					125				
15					GGG												432
	ser		Pro	Pro	Gly	Ala		Ser	Leu	Thr	Pro		Thr	Сув	Gly	Ser	
	mcc	130	amm				135					140					
					TTG												480
20	145	voħ	nea	TÄT	Leu	150	THE	Arg	HIS	ATA		Val	IIe	Pro	Val		
		CGG	GGC	GAC	AGC		GGG) NGC	CMC	CMC	155	C				160	514
					Ser							C					514
			1		165	1119	GLY	Der	neu	170	Ser						
25										1,0							
	SEQ ID	NO:4	19											-			
	SEQUEN	CE LE	engti	1: 52	23 ba	se p	pairs	5									
	SEQUEN	CE TY	PE:	nucl	leic	acio	i										
30	STRAND	EDNES	SS: d	loub]	e			•									
	TOPOLO	3Y:]	linea	ır													
	ANTI-SI	ense:	No.														
	ORIGINA																
35	ORGANIS																
	IMMEDIA			IMEN	ITAL	SOUF	RCE										
	CLONE:	026-	-3														
40	3 mg																
40					GCA												48
		Thr	Trp	GIY	Ala	Glu	Thr	Ala	Ala		Gly	yab	Ile	Ile		Gly	
	1 (ma	CCC	C III III	maa	5	005	100	~~~		10					15		
45					GCC												96
70	Dea	110	*aI	20	Ala	чтд	Arg	атА		GIU	тел	тел	ьeu		Pro	Ala	
				20					25					30			

	GAT	AGT	TTT	GAC	GGG	CAG	GGG	TGG	CGA	CTC	CTT	GCG	ССТ	ATC	ACG	GCC	144
	Asp	Ser	Phe	Asp	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	
			35					40					45				
5	TAC	TCC	CAG	CAA	ACG	CGG	GGC	CTG	CTT	GGT	TGC	ATC	ATC	ACT	AGC	CTT	192
	Tyr	Ser	Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Сув	Ile	Ile	Thr	Ser	Leu	
		50					55					60					
	ACG	GGC	CGG	GAT	AAA	AAC	CAG	GTC	GAG	GGG	GAG	GTT	CAA	GTG	GTC	TCT	240
10	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	
	65					70					75					80	
			ACA														288
	Thr	Ala	Thr	Gln	Ser	Phe	Leu	Ala	Thr	Cys	Val	Asn	Gly	Val	Cys	Trp	
15					85					90					95		
	ACT	GTT	TTC	CAC	GGT	GCC	GGC	TCG	AAG	ACC	TTA	GCC	GGC	CCA	AAA	GGC	336
	Thr	Val	Phe		Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	
				100					105					110			
20	CCA	ATC	ACC	CAA	ATG	TAC	ACC	TAA	GTG	GAT	CAG	GAC	CTC	GTC	GGT	TGG	384
	Pro	He	Thr	Gln	Met	Tyr	Thr		Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	
	maa	~~~	115					120					125				
			CCC														432
25	ser	ATA	Pro	Pro	Gly	Ala		Ser	Leu	Thr	Pro	Сув	Thr	Cys	Gly	Ser	
	maa	130					135					140					
			CTT														480
	ser	Asp	Leu	Tyr	Leu		Thr	Arg	His	Ala	Asp	Val	Ile	Pro	Val	His	
30	145	000				150					155					160	
			GGC												A		523
	wrd	arg	Gly	Asp		Arg	Gly	Ser	Leu		Ser	Pro	Gly	Pro			
					165					170							

SEQ ID NO:50

SEQUENCE LENGTH: 921 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

40 TOPOLOGY: linear

ANTI-SENSE: No

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

45 IMMEDIATE EXPERIMENTAL SOURCE

50

CLONE: N23-1

_	CTG	CTG	TCG	CCC	GGG	CCC	ATC	TCT	TAC	TTG	AAG	GGT	TCC	TCG	GGT	GGT	48
5	Leu	Leu	Ser	Pro	Gly	Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	Gly	
	1				5					10		•			15		
		CTG															96
10	Pro	Leu	Pro	Cys	Pro	Ser	Gly	Arg	Val	Val	Gly	Ile	Phe	Arg	Ala	Ala	
10				20					25					30			
		TGC															144
	Val	Cys		Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Val	Pro	Val	Glu	
15			35					40					45				
75		ATG															192
	Ser	Met	Glu	Thr	Thr	Met		Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	Thr	
		50					55					60					
20		CCG															240
20		Pro	ATA	Val	Pro		Thr	Phe	Gln	Val		His	Leu	His	Ala		
	65 200	ccc	N.C.C	cco		70	100				75					80	
		GGC															288
25	4111	Gly	Ser	GIY	85 85	Set	Inr	Arg	val	90	AIA	ATA.	туr	AIA		GIn	
	GGG	TAC	AAG	GTA	_	ርጥር	ርሞር	ממ	CCG		Cara	CCB	ccc	a Cm	95 mmc	ccc	226
		Tyr															336
	4	-1-	-1-	100			204	12011	105	Der	Val	AIG	AIG	110	ned	GIY	
30	TTT	GGG	GCG		ATG	TCC	AAG	GCA		GGT	GTT	GAC	CCT		ልጥሮ) AGD	384
		Gly															304
		_	115	_			•	120		2			125			9	
	ACT	GGG	GTG	AGG	ACC	ATC	ACC	ACG	GGC	GCT	CCC	ATC	ACG	TAC	TCC	ACC	432
35		Gly															
		130					135		_			140		_			
	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	GGT	GCC	TAT	GAC	480
	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Сув	Ser	Gly	Gly	Ala	Tyr	Asp	
40	145					150					155					160	
	ATC	ATA	ATA	T GT	GAT	GAG	TGT	CAT	TCA	ACT	GAC	TCG	ACT	TCC	ATC	TTG	528
	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	Asp	Ser.	Thr	Ser	Ile	Leu	
					165					170					175		
45	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG	GAG	ACG	GCT	GGA	GCG	CGC	CTT	576
	Gly	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	Leu	

50

				100													
				180					185					190			
			CTC														624
_	Val	Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	Thr	Val	Pro	His	
5			195					200					205				
			ATT														672
	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Asn	Thr	Gly	Glu	Ile	Pro	Phe	
		210					215					220					
10	TAT	GGC	AAG	GCC	ATC	CCC	CTC	GAG	GCC	ATC	AAG	GGG	GGG	AGG	CAT	CTC	720
			Lys														
	225					230					235		•			240	
	ATT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	GCG	AAG	CTG	768
15			Сув														
					245					250					255		
	TCG	GCC	CTC	GGA	GTC	AAC	GCT	GTA	GCA	TAT	TAC	CGG	GGT	CTT	GAT	GTG	816
			Leu														
20				260					265	•	-	-	-	270			
	TCC	ATC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACA	GAC	GCT	864
			Ile														•••
			275				- 4	280					285		1.Up	1114	
25	CTA	ATG	ACG	GGC	TAT	ACC	GGT	GAC	սար	GAC	ጥሮር	GTG.		GAC	ጥርር	አአሮ	912
			Thr														312
		290		•	-		295				001	300		p	Cys	Non	
	ACA	TGT	GTC									300					921
30	Thr	Cys	Val														341
30	305				-								•	•			
	SEQ ID	NO:5	i 1														
	CRUITEN																

SEQUENCE LENGTH: 921 base pairs
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ANTI-SENSE: No
ORIGINAL SOURCE

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE

CLONE: N23-2

50

	CTG	CTG	TCG	CCC	GGG	ccc	ATC	TCC	TAC	CTG	AAG	GGT	TCC	TCG	GGT	GGT	48
	Leu	Leu	Ser	Pro	Gly	Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	Gly	
•	1				5					10					15	_	
5	CCG	CTG	CTT	TGC	CCC	TCG	GGC	CAT	GTT	GTG	GGC	ATC	TTC	CGG	GCT	GCT	96
	Pro	Leu	Leu	Cys	Pro	Ser	Gly	His	Val	Val	Gly	Ile	Phe	Arg	Ala	Ala	
				20					25					30			
10	GTG	TGC	ACC	CGG	GGG	GTT	GCG	AAG	GCG	GTA	GAC	TTT	GTG	CCC	GTT	GAG	144
10	Val	Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Val	Pro	Val	Glu	
			35					40					45				
	TCT	ATG	GAA	ACC	ACT	ATG	CGG	TCT	CCG	GTC	TTC	ACG	GAT	AAC	TCA	ACC	192
4-	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	Thr	
15		50					55					60					
					CCG												240
	Pro	Pro	Ala	Val	Pro	Gln	Ser	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	
00	65					70					75					80	
20					AAA												288
	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Lys	Val	Pro	Ala	Ala	Tyr	Ala	Ala	Gln	
					85					90					95		
25					CTC												336
25	Gly	Tyr	Lys		Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	
				100					105					110			
					ATG												384
- 30	Phe	GIA		Tyr	Met	Ser	Lys		His	Gly	Val	Asp		Asn	Ile	Arg	
30	3.00		115					120					125		-		
					ACC												432
	THE	130	vai	Arg	Thr	тте		Thr	GIY	Ala	Pro		Thr	Tyr	Ser	Thr	
35	ሞልር		220	mme	CMC	ccc	135	00m	000	mam		140					
00					CTC												480
	145	GLY	шуз	rne	Leu	150	АБР	GIY	GTĀ	Сув		GIĀ	GTĀ	ATA	туг		
		ፈጥ ል	ልጥል	ጥርረጥ	GAT		anca an	Cam	mc a	» Cm	155	maa	> Cm	maa	> ma	160	500
40					Asp												528
40			-10	Cys	165	GIU	Cys	mre	Ser	170	мър	Ser	THE	ser		ьеп	
	GGC	ΔͲͲ	CCT	ACA	GTC	CTC	CNC	CAA	ccc		» CC	ccm	CCI	COC	175	omm.	F76
					Val												576
45	1		1	180	741	-eu	vob	GIH	185	GIU	TILE	HIG	атĀ	190	Arg	ьeu	
70	GTC	GTG	СТС		ACC	CCT	ACG.	CC ም		GGA	ምሮሮ	CMC	NCC		ccc	C N m	624
							AUG			JUN	100	GIC	ACC	GT.G	CCG	CAT	624

	Val	Val	Leu 195		Thr	Ala	Thr	Pro		Gly	Ser	Val	Thr 205		. Pro	His	
	CCT	' AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ልሮሞ	GGA			CCC	TTC	672
5	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Asn	Thr	Glv	Glu	Tle	Pro	Phe	072
		210					215					220				FIIG	
	TAT	GGC	AAG	GCC	ATC	CCC	CTC	GAG	GCC	ATC	AAG	GGG	GGG	AGG	CAT	CTC	720
	Tyr	Gly	Lys	Ala	Ile	Pro	Leu	Glu	Ala	Ile	Lys	Gly	Gly	Arq	His	Leu	
10	225					230					235	-	-	_		240	
	ACT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	GCG	AAG	CTG	768
	Thr	Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	
					245					250					255		
15	TCG	GCC	CTC	GGA	GTC	AAC	GCT	GTA	GCA	TAC	TAC	CGG	GGT	CTT	GAT	GTG	816
	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	Val	
				260					265			,		270			
	TCC	GTC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACT	GAC	GCT	864
20	Ser	Val	Ile	Pro	Thr	Ser	Gly	Asp	Val	Val	Val	Val	Ala	Thr	Asp	Ala	
			275					280					285				
	CTA	ATG	ACG	GGC	TAT	ACC	GGT	GAC	TTT	GAC	TCA	GTG	ATC	GAC	TGC	AAC	912
	Leu	Met	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	
25		290					295					300					
		TGT															921
		Cys	Val														
	305																
30	450		-		ř								•				•
	SEQ ID												•				
	SEQUENC							;									
Ω	SEQUENC					acid	Ĺ										
35	STRANDE				e												
	TOPOLOG			r													
	ANTI-SE																
	ORIGINA				_												
40	ORGANIS																
	IMMEDIA CLONE:			TWEN	TAL	SOUR	CE										
	CHONE:	1423-	3														
45	CTG	CTG	TCG	CCC -	ממני י	רככ	አ ጥረግ ፡	መረጠ ፡	ma.a	mm~	220		ma-				
70	CTG Leu	Leu	Ser	Pro	Clu (Dro	Tla	CO-	TAC	TTG	AAG T	GGC G1	TCC	TCG	GGT	GGT	48
	Leu			- 10	CIY .	LIO	TTG	ser	TÄL	ьeи	тĀв	GΤĀ	ser	ser	Gly	Gly	

	1				5					10					15		
	CCG	CTG	CTT	TGC	CCC	TCG	GGC	CAT	GTT	GTG	GGC	ATC	TTC	CGG		GCC	96
					Pro												30
5				20			_		25		_			30			
	GTG	TGC	ACC	CGG	GGG	GTT	GCG	AAG	GCG	GTG	GAC	TTT	GTG	CCC	GTT	GAG	144
					Gly												
	•		35					40			_		45				
10	TCT	ATG	GAA	ACC	ACC	ATG	CGG	TCT	CCG	GTC	TTC	GCG	GAT	AAC	TCA	ACC	192
	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Ala	Asp	Asn	Ser	Thr	
		50					55		•			60					
	CCC	CCG	GCC	GTA	CCG	CAG	ACA	TTC	CAA	GTG	GCC	CAC	CTA	CAC	GCT	CCC	240
15	Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	
	65					70					75					80	
	ACT	GGC	AGC	GGC	AAA	AGC	ACC	AGG	GTG	CCG	GCT	GCG	TAT	GCG	GCC	CAA	288
	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Arg	Val	Pro	Ala	Ala	Tyr	Ala	Ala	Ġln	
20					85					90					95		
					CTC												336
	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	
05				100					105					110			
25					ATG												384
	Phe	Gly		Tyr	Met	Ser	Lys	Ala	His	Gly	Val	qaA	Pro	Asn	Ile	Arg	
			115					120					125				
00					ACC												432
30	Thr		Val	Arg	Thr	Ile		Thr	Gly	Ala	Pro	Val	Thr	Tyr	Ser	Thr	
	mao	130					135					140		•			
					CTC												480
35	145	GIY	гув	Pne	Leu		Asp	GLY	GLY	Сув		Gly	Gly	Ala	Tyr	Asp	
33		מידות	እመጸ	m/cm	C N M	150	man	~~			155					160	
					GAT												528
	116	TTE	TTE	cys	Asp	GIU	Сув	Hls	ser		Asp	Ser	Thr	Ser		Leu	
40	ccc	አጥጠ	CCM	202	165	ama	~~~			170					175		
40					GTC												576
	GLY	116	GLY	180	Val	ren	Asp	GIN		Glu	Thr	Ala	Gly		Arg	Leu	
	GTC	GTG	ርሞር		N.C.C	CCE	3.00	com	185	a a>				190			
45					ACC												624
45	- 44		195	wata	Thr	via			LI.O	атХ	ser	val		val	Pro	His	
			490					200					205				

	ርር ሞ	ልልጥ	ሃ ብուն	GAG	GAG	CITIC	ccc	mmc	mcc	220	3.00	001	a.a	3 000			
					Glu												672
		210		014	GIU	Vai	215	nea	per	Man	THE	220	GIU	116	Pro	Pne	
5	TAT		AAG	GCC	ATC	CCC		GAG	GCC	ልጥሮ	AAG		GGG	AGG	ሮአመ	CMC	720
					Ile												120
	225	- 4				230					235	071	913	мy	1170	240	
	ATT	TTC	TGC	САТ	TCC		AAG	AAA	TGT	GAC		CTC	GCT	GCG	AAG		768
10					Ser												700
			•		245	•	-		-1-	250					255	Dea	
	TCG	GCC	CTC	GGA	GTC	AAT	GCT	GTA	GCA		TAC	CGG	GGT	СТТ		GTG	816
	Ser																0_0
15				260					265	•	•	3	2	270			
	TCC	ATC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA-	ACA	GAC	GCT	864
	Ser																
			275					280					285		-		
20	CTA .	ATG	ACG	GGC	TAT	ACC	GGT	GAC	TTT	GAC	TCG	GTG	ATC	GAC	TGT	AAC	912
	Leu :	Met	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Сув	Asn	
		290					295					300					
	ACA	TGT	GTC														921
25	Thr	Сув	Val														
	305																
	SEQ ID												•				
30	SEQUENC					_											
	SEQUENC					acio	ì										
	STRANDE				e												
	TOPOLOG			ır													
35	ANTI-SE																
	ORIGINA						_										
	ORGANIS									-							
	CLONE:			LIMEN	TAL	SOUR	CE										
40	CHORE:	MT0-															
	GGC :	TAT	ACC	GGC	GAC	ጥጥር	CDC	ጥሮ አ	CTYC	አጥሮ	CNC	mcc	አአጣ	202	anca m	CMC	40
	Gly																48
	1	-1-		~-1	7.5 p	- 116	.wp	JUL	4 G I	10	veħ	Cyb	VOII	THE	Cys 15	AGT	
45	ACC (CAA	ACA	GTC	_	TTC	AGC	ጥ ፕር	GAC		ልሮሞ	ጥጥረግ	ACC	ልጥሮ		ACG	96
													-100	-110	JAG	2100	J 0

	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
				20					25					30			
			GTA														144
5	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
			35					40					45				
			AGG														192
	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
10		50					55					60					
			TCA														240
		Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	Glu	Cys	Tyr	Asp	
	65					70					75					80	
15			TGT														288
	Ala	Gly	Сув	Ala		Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Ser	Val	Arg	
					85					90					95		
			GCT														336
20	Leu	Arg	Ala		Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Сув	Gln	Asp	His	
				100					105					110			
			TTC														384
	Leu	Glu	Phe	Trp	Glu	Ser	Val		Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
25			115					120					125				
			TTG -												•		432
	H1S		Leu	Ser	Gln	Thr		Gln	Ala	Gly	Asp		Phe	Pro	Tyr	Leu	
	am.	130	m. a	~~~			135					140	•				
30			TAC														480
	145	AIA	Tyr	GIN	ATA		val	Сув	Ala	Arg		Lys	Ala	Pro	Pro		
		יייי	Cam	CNC	3 MC	150		mam	-		155					160	
			GAT														528
35	Sel	пр	Asp	GIII		тгр	тАв	cys	Leu		Arg	Leu	Lys	Pro		Leu	
	CAC	ccc	CCA	3.00	165	ama	mmo	m > m		170		·			175		
			CCA														576
	ure.	GTĀ	Pro		Pro	ren	ren	тХт		ren	GLY	Ala	Val		Asn	Glu	
40	Citran	እሮሮ	Cmm	180	a.a	000			185					190			
			CTT													TC	623
	AGT	THE	Leu	THE	uls	PTO	TTE		гав	туr	TTO	Met		Cys	Met		
			195					200					205				

45 SEQ ID NO:54

55

SEQUENCE LENGTH: 623 base pairs
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
ANTI-SENSE: NO
ORIGINAL SOURCE
ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE
CLONE: N16-2

	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCA	GTG	ATC	GAC	TGC	AAC	ACA	TGT	GTC	48
15	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	Thr	Сув	Val	
	1				5					10					15		
	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	CCT	ACT	TTC	ACC	ATC	GAG	ACG	96
	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
20				20					25					30			
					CAA												144
	Thr	Thr		Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
			35					40					45				
25					AGG												192
	Thr		Arg	GlĀ	Arg	Gly		Ile	Tyr	Arg	Phe		Thr	Pro	Gly	Glu	
	ccc	50	ma.	000	200		55					60					
					ATG												240
30 .	65		ser	GTA	Met		Asp	ser	Ser	Val		Cys	Glu	Cys	Tyr		
		cer	mvm	CCM	maa	70	63.6	ama			75			<u>.</u>		80	
					TGG												288
		Gry	Cys	та	Trp 85	TYL	GIU	Leu	Thr		AIa	Glu	Thr	Ser		Arg	
35	TTG	CGG	ር ርጥ	ሞልሮ		አ አጥ	አሮአ	ccm	ccc	90	000	oma	maa	~~~	95		
					CTA Leu												336
		9		100	neu	non	1111	FIO	105	nea	PIO	var	Сув	110	Asp	HIS	
	CTG	GAG	TTC		GAG	AGC	ርጥሮ	ጥጥር		ccc	CTC	N.C.C	CNC		CNE	000	204
40					Glu												384
			115			002	,,,,	120		GIY	пеа	IIII.	125	116	Авр	Ala	
	CAC	TTC	TTG	TCC	CAG	ACC	AAA		GCA	GGA	GAC	מממ		ccc	መአረግ	CITIC	432
					Gln												432
45		130					135	- 		J-7	-105	140	- MC	.10	TYL	⊔ ∈u	
												~ Z U					

50

																CCA	480
	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Lys	Ala	Pro	Pro	Pro	
	145					150					155					160	
5			GAT														528
	Ser	Trp	Asp	Gln	Met	Trp	Lys	Суs	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	
					165					170					175		
			CCA														576
10	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	
				180					185					190			
			CTC													TC	623
	Val	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	Tyr	Ile	Met	Thr	Сув	Met		
15			195					200					205				
	SEQ ID																
	SEQUEN					_		3									
20	SEQUEN					acio	İ						•				
	STRAND				le												
	TOPOLO			ar													
	ANTI-S			_												•	
25	ORIGIN																
	ORGANI																
	IMMEDIA			RIME	TAL	SOU	RCE										
	CLONE:	NTP.	-3														
30	000	.m.s.m	200														
			ACC														48
	GIY 1	TYL	Thr	GIY		Phe	Asp	Ser	Val		Asp	Cys	Asn	Thr		Val	
		CAA	202	CMC	5	mma	100		a. a	10					15		
35			ACA														96
	1111	GIII	Thr	20	Asp	Pne	ser	ren		Pro	Thr	Phe	Thr		Glu	Thr	
	»CC	NCC.	Cmx		C2.3	C> m	000	ama.	25					30			
			GTA														144
40	1111	1111	Val	PLO	GIII	Asp	ALA		ser	Arg	ser	GIn		Arg	Gly	Arg	
	ልሮመ	CCm	35 ACC	ccc	202	cca	000	40	mr c			a	45				
			AGG														192
	THE		Arg	стĀ	Arg	GТĀ		тте	туr	Arg	Phe		Thr	Pro	Gly	Glu	
AE	000	50	mer »	000	1 60	mma	55					60					
4 5	CGG	CCC	TCA	GGC	ATG	TTC	GAT	TCT	TCG	GTC	CTG	TGT	GAA	TGT	TAT	GAC	240

	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Сув	Glu	Cys	Tyr	Asp	
	65					70					75					80	
	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACG	TCC	GCC	GAG	ACC	TCG	GTT	AGG	288
5	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	Ser	Ala	Glu	Thr	Ser	Val	Arg	
					85					90					95		
	TTG	CGG	GCT	TAC	CTA	AAC	ACA	CCT	GGG	CTG	CCC	GTC	TGC	CAG	GAC	CAT	336
	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys	Gln	Asp	His	
10				100					105					110			
	CTG	GAG	TTC	TGG	GAG	AGC	GTC	TTC	ACC	GGC	CTC	ACC	CAC	ATA	GAT	GCC	384
	Leu	Glu	Phe	Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
			115					120					125				
15	CAC	TTC	TTG	TCC	CAG	ACT	AAA	CAG	GCA	GGA	GAC	AAC	TTC	CCC	TAC	CTG	432
	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	
		130					135					140	•				
	GTA	GCA	TAC	CAG	GCT	ACA	GTG	TGC	GCC	AGG	GCC	AAG	GCT	CCA	CCT	CCA	480
20	Val	Ala	Tyr	Gln	Ala	Thr	Val	Сув	Ala	Arg	Ala	Lys	Ala	Pro	Pro	Pro	
	145					150					155					160	
	TCG	TGG	GAT	CAG	ATG	TGG	AAG	TGT	CTC	ATA	CGG	CTG	AAG	CCT	ACG	CTA	528
	Ser	Trp	Asp	Gln	Met	\mathtt{Trp}	Lys	Cys	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	
25					165					170					175		
	CAC	GGG	CCA	ACG	CCC	CTG	TTG	CAT	AGG	TTA	GGA	GCC	GTT	CAG	AAC	AAG	576
	His	Gly	Pro	Thr	Pro	Leu	Leu	His	Arg	Leu	Gly	Ala	Val	Gln	Asn	Lys	
				180					185					190		•	
30						CCC										TC	623
	Val	Ala	Leu	Thr	His	Pro	Ile	Thr	Lys	Tyr	Ile	Met	Thr	Cys	Met		
			195					200					205				

SEQ ID NO:56

SEQUENCE LENGTH: 1280 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: No

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

45 CLONE: MX25026A-1

50

35

	TGT	GCC	TGG	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAA	GCI	GAG	GCC	GCC	48
	Cys	Ala	Trp	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	
	1				5					10					15		
5	TTG	GAG	AAC	CTG	GTG	GTC	CTC	AAT	GCA	GCA	TCC	ATG	GCG	GGA	GCG	CAT	96
	Leu	Glu	Asn	Leu	Val	Val	Leu	Asn	Ala	Ala	Ser	Met	Ala	Gly	Ala	His	
				20					25					30			
	GGC	ATC	CTC	TCT	TTC	CTT	GTG	TTC	TTC	TGT	GCC	GCC	TGG	TAC	ATC	AAA	144
10	Gly	Ile	Leu	Ser	Phe	Leu	Val	Phe	Phe	Сув	Ala	Ala	Trp	Tyr	Ile	Lys	
			35					40					45				
	GGC	AGG	CTG	GTC	CCT	GGG	GCG	GCA	TAC	GCT	TTC	TAT	GGC	GTA	TGG	CCG	192
	Gly		Leu	Val	Pro	Gly	Ala	Ala	Tyr	Ala	Phe	Tyr	Gly	Val	Trp	Pro	
15		50					55					60					
	CTG	CTC	CTG	CTC	TTG	ATG	GCG	CTA	CCC	GCA	CGG	GCG	TAC	GCC	ATG	GAC	240
		Leu	Leu	Leu	Leu	Met	Ala	Leu	Pro	Ala	Arg	Ala	Tyr	Ala	Met	Asp	
	65					70					75					80	
20												TTT					288
	Arg	Glu	Met	Ala		Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	Leu	Val	
	oma	mmo			85					90					95		
	CTC	TIG	ACC	TTG	TCA	CCA	TAC	TAC	AAA	GTG	TTC	CTC	GCT	AAG	CTC	ATA	336
25	neu	rea	Thr		ser	Pro	Tyr	Tyr		Val	Phe	Leu	Ala	Lys	Leu	Ile	
	TYCC	TICC.	mmc	100		ama			105					110			
	Trn	Ψrn 100	Lon	CIA	TAT	CIC	ATC	ACC	AGG	GCC	GAG	GCG	CAC	TTG	CAA	GTG	384
	115	11p	115	GIII	TÄT	ren	me		Arg	Ala	Glu	Ala			Gln	Val	
30	TGG	ATC		ררר	ርሞር	ממ	Cmm	120	ccc	cca	000	GAT	125				
	Tro	Ile	Pro	Pro	T.An	Acn	Ual	7~~	C1	GGC	CGC	Asp	GCC	ATC	ATC	CTT	432
		130				*1011	135	ALG	GIY	GIY	Arg		ALG	TTE	TTE	Leu	
	CTC		TGT	GCG	GTC	CAC		GAG	ርሞር	አጥሮ	mmm	140 GAC	3.00	300			400
35	Leu	Thr	Cvs	Ala	Val	His	Pro	Glu	Len	Tla	Dha	Asp	TIO	Mb-	AAG	CTC	480
	145		-			150		- Cau	DC4	116	155	veħ	116	THE	гув		
	TTG	CTC	GCC	ATA	CTC		CCG	СТС	АПС	СТА		CAG	CCT	ccc	COS	160	500
	Leu	Leu	Ala	Ile	Leu	Glv	Pro	Leu	Met.	Val	Len	Gln	Ala	Clar	LON	Mb~	528
40					165					170	200	0111	niu	GIY	175	THE	
	CAA	ATG	CCG	TAC	TTT	GTG	CGT	GCT	CAA		CTC	ÀΤΤ	CGT	2ንሞል		ልጥር	576
	Gln	Met	Pro	Tyr	Phe	Val	Arg	Ala	Gln	Glv	Leu	Ile	Ara	Met	Cvs	Met	370
				180			_		185	-			5	190	-1-		
45	TTG	GTG	CGG .	AAA	GCC	GCT	GGG	GGT	CAT	TAT	GTC	CAG	ATG		CTC	ATG	624
															0		027

	Leu	Val			Ala	Ala	Gly	Gly	His	Туг	. Val	Gln	Met	Ala	Leu	Met	
			195					200			•		205				
	AAG	CTG	GCT	GCA	CTG	ACA	GGT	ACG	TAC	GTT	TAT	GAC	CAT	CTT	ACT	CCA	672
5	гйг			Ala	Leu	Thr			Туг	Val	Tyr	qaA	His	Leu	Thr	Pro	
	CMC	210					215					220					
	Tan	CAG	GAC	TGG	GCC	CAC	GCG	GGC	CTA	CGA	GAC	CTT	GCG	GTA	GCA	GTT	720
	225	GIII	Asp	Trp	ALA			GLY	Leu	Arg		Leu	Ala	Val	Ala	Val	
10		ccc	CMM	666	mma	230					235					240	
	GAG	Bro	GII	GCC 3.1.	TTC	TCT	GAT	ATG	GAG	ACT	AAG	ATC	ATC	ACC	TGG	GGG	768
	Giu	FIO	vaı	AId		ser	Asp	Met	Glu			Ile	Ile	Thr	Trp	Gly	
	CCA	GNC	».cm	coc	245	mam				250					255		
15															GTC		816
	NIG	nsp	THE	260	ATG	Cys	GIY	Авр		He	Leu	Gly	Leu		Val	Ser	
	ccc	CGG	እርር		330	CAC	3.003	0ma	265					270			
	Ala	724	Ava	C1	AAC	GAG	ATA	CTC	CTC	GGA	CCG	GCC	GAT	AGT	TTT	GAC	864
20	111U	мy	275	GIŞ	ABII	GIU	TIE		ьеи	GIY	Pro	Ala		Ser	Phe	Asp	
	GGG	CAG		ጥርር	CCA	CMC	Cmm	280	aam				285				
	Glv	Gln	Glv	Trn) TOA	Ton	LIT	31-	CCT	ATC	ACG	GCC	TAC	TCC	CAG	CAG	912
	ردن	290	GLY	TTP	мц	nea	295	AIG	Pro	TIE	Thr		Tyr	Ser	Gln	Gln	
25	ACG		GGC	ርጥር	C. dreft	CCT		አመር	አመሮ	a Cm	300	300			CGG		
	Thr	Arg	Glv	Leu	Len	Clv	Cac	Tla	AIC	Mb~	AGC	CTT	ACG	GGC	Arg	GAT	960
	305	3	1		204	310	cys	116		THE	315	Leu	THE	GIĀ	Arg		
	AAG	AAC	CAG	GTC	GAG		GAG	CTT	CAA	CTC		መረመ	»CC	CCX	ACA	320	1000
30	Lys	Asn	Gln	Val	Glu	Glv	Glu	Val	Gln	Val	Val	Ser	Thr.	Ala	Thr	CAA	1008
					325	1			411	330	V 41	Ser	1117	MIG	335	GIII	
	TCT	TTC	CTG	GCG	ACC	TGT	GTC	AAC	GGC		TGC	ጥርር	ልሮሞ	Com	TTC	CAC	1056
															Phe		1030
35				340		-			345		-1 -			350	THE	117.0	
	GGC	GCC	GGC	TCG	AAG	ACC	TTA	GCC	GGC	CCA	AAA	GGC	CCA		ACC	CAA	1104
	Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Glv	Pro	Lvs	Glv	Pro	Tla	Thr	Gla	1104
			355		-			360	2		-1-	0-1	365	110	11	GIII	
40	ATG	TAC	ACC	AAT	GTA	GAT	CAG	GAC	CTC	GTC	GGC	TGG		GCG	ccc	CCC	1152
															Pro		11,72
		370				-	375	•				380			- 10	- 10	
	CGG	GCG	CGT	TCC	TTG	ACA	CCT	TGC	ACC	TGC	GGC		TCG	GAC	CTT	TAT	1200
45	Arg	Ala	Arg	Ser	Leu	Thr	Pro	Cys	Thr	Сув	Glv	Ser	Ser	Asn	Len	ጥህን ጉህን	-200
			-					•		- 2 -	1				20 u	-1-	

	385					390					395					400	
			ACG														1248
	Leu	Val	Thr	Arg	His	Ala	Asp	Val	Ile	Pro	Val	His	Arg	Arg	Gly	Asp	
5					405					410					415		
			GGG								ΑT						1280
	Ser	Arg	Gly		Leu	Leu	Ser	Pro	Gly	Pro			•				
				420					425								
10														*			
	SEQ ID		- '														
	SEQUEN						_	rs									
	SEQUEN					acio	i										
15	STRAND				le												
	TOPOLO			ar													
	ANTI-SI																
	ORIGINA																
20	ORGANIS		_														
	IMMEDIA				NTAL	SOU	RCE										
	CLONE:	MX 2	50261	3-1													
	יועכיתי	ccc	mcc	mmc	mcc.	3 000	3.000	ama	ama								
25			TGG														48
	Cys 1	VIG	Trp	neu	11.p	Met	met	rea	ren		ATA	GIN	ATA	GIu		Ala	
	_	GAG	AAC	CTC	_	CTC	CMC	220	CCX	10	maa	3.000	000		15		•
			Asn														96
30				20		•	Deu	Aon	25	AIG	Ser	Met	AIG	30	AIA	nıs	
	GGC	ATC	CTC		ምጥር	СФФ	CTC	ጥጥር		ጥርጥ	GCC	GCC	mcc		a mc	222	144
			Leu														144
	-		35					40		C1 D	2124	ALU	45	TYL	116	nys	
3 5	GGC	AGG	CTG	GTC	ССТ	GGG	GCG	GCA	TAC	GCT	ттс	тат		СТА	TCC	CCG	192
			Leu														172
	:-	50				_	55		-1			60	011		P	110	
	CTG	CTC	CTG	CTC	TTG	ATG		СТА	CCC	GCA	CGG		TAC	GCC	атс	GAC	240
40			Leu														210
	65					70					75					80	
	CGG	GAG	ATG	GCT	GCA		TGC	GGA	GGC	GCG		TTT	GTA	GGT	CTG		288
			Met														
45	-				85		-	-	-4	90				1	95		
					-					_ •					,,		

5

	CTC	TTG	ACC	TTG	TCA	CCA	TAC	TAC	AAA	GTG	TTC	CTC	GCT	AAG	CTC	ATA	336
	Leu	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Val	Phe	Leu	Ala	Lys	Leu	Ile	
				100					105					110			
5																GTG	384
	Trp	Trp	Leu	Gln	Tyr	Leu	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	Gln	Val	
			115					120					125				
			CCC														432
10	Trp		Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Ile	Ile	Leu	
		130					135					140					
			TGT														480
		Thr	Cys	Ala	Val	His	Pro	Glu	Leu	Ile	Phe	Asp	Ile	Thr	Lys	Leu	
15	145					150					155					160	
			GCC														528
	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Gly	Leu	Thr	
					165					170					175		
20			CCG														576
	Gin	Met	Pro		Phe	Val	Arg	Ala	Gln	Gly	Leu	Ile	Arg	Met	Cys	Met	
	mma			180					185					190			
			CGG														624
25	ren	vaı	Arg	Lys	Ala	Ala	Gly		His	Tyr	Val	Gln	Met	Ala	Leu	Met	
		ama.	195					200					205				
			GCT														672
	гув		Ala	Ата	Leu	Thr		Thr	Tyr	Val	Tyr		His	Leu	Thr	Pro	
30	CMC	210	222				215					220					
			GAC														720
	225	GIN	Asp	Trp	ALA		Ala	GIY	Leu	Arg		Leu	Ala	Val	Ala		
		ccc	CMM	CCC	mmc	230	a				235					240	
35			GTT														768
	GIU	FIO	Val	мта	245	ser	Asp	Met	GIU		гÃ2	тте	ITe	Thr	_	GIĀ	
	GCA	CAC	ACG	GCG		mcm	ccc	CNC	3 mg	250	maa	aam.		222	255		
			ACG														816
40			Thr	260	TIG	Cys	GTÅ	Asp		тте	Ser	GIĀ	rea		vaı	ser	
	GCC	ርር እ	AGG) CC	CNC	COLC	com	265 mmc	cca	000	cca	a.m	270		~~~	064
			AGG														864
		9	Arg 275	31y	ALY.	GIU	TIGIT		тел	GTÅ	Pro	WIG		ser	rue	Asp	
45	GGG	CAG		ጥርረር	ርር እ	CIPC	Cum	280	a a a a	» ma	200	000	285	mc~	a	03.0	010
			GGG	193	CGM	CIL	CIT	GCG	CCT	ATC	AUG	GCC	TAC	TCC	CAG	CAG	912

	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ser	Gln	Gln	
		290					295					300					
	ACG	CGG	GGC	CTG	CTT	GGT	TGC	ATC	ATC	ACT	AGC	CTT	ACG	GGC	CGG	GAT	960
5	Thr	Arg	Gly	Leu	Leu	Gly	Сув	Ile	Ile	Thr	Ser	Leu	Thr	Gly	Arg	Asp	
	305					310					315					320	
	AAG	AAC	CAG	GTC	GAG	GGG	GAG	GTT	CAA	GTG	GTC	TCT	ACC	GCA	ACA	CAA	1008
	Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	Thr	Ala	Thr	Gln	
10					325					330					335		
			CTG														1056
	Ser	Phe	Leu	Ala	Thr	Cys	Val	Asn	Gly	Val	Сув	Trp	Thr	Val	Phe	His	
				340					345			٠		350			
15			GGC														1104
	Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	Pro	Ile	Thr	Gln	
			355					360					365				
			ACC														1152
20	Met		Thr	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	Ser	Ala	Pro	Pro	
		370					375					380					
		GCG	CGT				CCT					AGC					1200
	Arg	GCG	CGT Arg				CCT					AGC					1200
25	Arg 385	GCG Ala	Arg	Ser	Leu	Thr 390	CCT Pro	Сув	Thr	Суѕ	Gly 395	AGC Ser	Ser	Asp	Leu	Tyr 400	1200
25	Arg 385 TTG	GCG Ala GTC	Arg ACG	Ser AGG	Leu CAT	Thr 390 GCT	CCT Pro	Cys GTC	Thr ATT	Cys CCG	Gly 395 GTG	AGC Ser	Ser CGG	Asp CGG	Leu GGC	Tyr 400 GAC	1200
25	Arg 385 TTG	GCG Ala GTC	Arg	Ser AGG	Leu CAT His	Thr 390 GCT	CCT Pro	Cys GTC	Thr ATT	Cys CCG	Gly 395 GTG	AGC Ser	Ser CGG	Asp CGG	Leu GGC	Tyr 400 GAC	
25	Arg 385 TTG Leu	GCG Ala GTC Val	Arg ACG Thr	Ser AGG Arg	CAT His 405	Thr 390 GCT Ala	CCT Pro GAT Asp	Cys GTC Val	Thr ATT Ile	Cys CCG Pro 410	Gly 395 GTG Val	AGC Ser	Ser CGG	Asp CGG	Leu GGC	Tyr 400 GAC	
25	Arg 385 TTG Leu AGC	GCG Ala GTC Val	ACG Thr	Ser AGG Arg AGC	CAT His 405 CTC	Thr 390 GCT Ala CTC	CCT Pro GAT Asp	Cys GTC Val	Thr ATT Ile	Cys CCG Pro 410 CCC	Gly 395 GTG Val	AGC Ser	Ser CGG	Asp CGG	Leu GGC Gly	Tyr 400 GAC	
	Arg 385 TTG Leu AGC	GCG Ala GTC Val	Arg ACG Thr GGG Gly	Ser AGG Arg AGC	CAT His 405 CTC	Thr 390 GCT Ala CTC	CCT Pro GAT Asp	Cys GTC Val	Thr ATT Ile	Cys CCG Pro 410 CCC	Gly 395 GTG Val	AGC Ser	Ser CGG	Asp CGG	Leu GGC Gly	Tyr 400 GAC	1248

SEQ ID NO:58

SEQUENCE LENGTH: 1431 base pairs SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: NO
ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

45 CLONE: N16N15A-1

50

35

	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCA	GTG	ATC	GAC	TGC	AAC	ACA	TGT	GTC	48
	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	Thr	Cys	Val	
-	1				5					10					15		
5	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	CCT	ACT	TTC	ACC	ATC	GAG	ACG	96
	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
				20					25					30			
					CAA												144
10	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
			35					40					45				
					AGG												192
	Thr		Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
15		50					55					60					
					ATG												240
		Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Сув	Glu	Cys	Tyr	Asp	
	65					70					75					80	
20					TGG												288
	Ala	Gly	Суѕ	Ala	Trp	Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Ser	Val	Arg	
					85					90					95		
					CTA												336
25	Leu	Arg	Ala		Leu	Asn	Thr	Pro		Leu	Pro	Val	Сув	Gln	Asp	His	
	ama	~~~		100					105					110			
					GAG												384
	rea	GIU		Trp	Glu	ser	vaı		Thr	Gly	Leu	Thr		Ile	Asp	Ala	
30	CAC	መጥር	115	mcc	CNC	300		120	222				125				
					CAG												432
		130	nea	Ser	Gln	THE	цув 135	GIN	AIA	GIĀ	Asp		Pne	Pro	Tyr	Leu	
05	СТА		ጥልሮ	CAG	GCT	አሮአ		mcc	cca	100	000	140	o a m				
35					Ala												480
	145		-1-	0111	AIG	150	Val	cys	ALG	ALG	155	тЛя	AIA	PIO	Pro		
		ፕሮር	ርልጥ	CAG	ATG		AAC	መረመ	CITIC	nma.		CITIC		com		160	500
40					Met												528
40			·p		165		цуз	Cys	пеп	170	ALG	Leu	гуя	PIO		ren	
	CAC	GGG	CCA	ACG	CCC	ርሞር	ጥጥር	ጥልጥ	ACC		CCA	ccc	cmm	CNC	175	ara.	576
					Pro												576
45		2		180				-1-	185	⊒eu.	GTÄ	wra	val	190	ASN	GIU	
45	GTT	ACC	СТТ		CAC	CCC	בידה	ACC		ጥልሮ	ልጥሮ	ልሞር	AC 2		አመሮ	mac.	624
								-100	• 1111	INC	arro	HIG	AUA	TGC	AIG	100	624

			195					200			Ile		205				
	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	TGG	GTG	CTG	GTA	GGC	GGG	GTC	672
5	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	Val	Gly	Gly	Val	
		210					215					220					
	CTC	GCG	GCT	CTG	GCC	GCG	TAC	TGC	CTG	ACA	ACG	GGC	AGC	GTG	GTC	ATT	720
	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Сув	Leu	Thr	Thr	Gly	Ser	Val	Val	Ile	
10	225					230					235					240	
	GTG	GGC	AGG	ATC	ATC	TTG	TCC	GGG	AGG	CCG	GCC	GTT	ATT	CCC	GAC	AGG	768
	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	Val	Ile	Pro	Asp	Arg	
					245					250					255		
15	GAA	GTT	CTC	TAC	CAA	GAG	TTC	GAT	GAA	ATG	GAA	GAG	TGC	GCC	TCG	CAC	816
	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	Сув	Ala	Ser	His	
				260					265	•				270			
	CTC	CCT	TAC	ATC	GAA	CAA	GGA	ATG	CAG	CTC	GCC	GAG	CAA	TTC	AAG	CAG	864
20	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu.	Gln	Phe	Lys	Gln	
			275					280					285				
	AAG	GCG	CTC	GGT	TTG	CTG	CAA	ACA	GCC	ACC	AAG	CAA	GCG	GAG	GCT	GCT	912
	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	Gln	Ala	Glu	Ala	Ala	
25		290					295					300					
											CTT						960
		Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	Glu	Thr	Phe	Trp	Ala	
	305					310					315					320	
30											CAG						1008
	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	Tyr	Leu	Ala	Gly	Leu	
					325					330					335		
											TCA						1056
35	Ser	Thr	Leu		Gly	Asn	Pro	Ala	Ile	Ala	Ser	Leu	Met	Ala	Phe	Thr	
				340					345					350			
											TAT						1104
	Ala	Ser		Thr	Ser	Pro	Leu	Thr	Thr	Gln	Tyr	Thr	Leu	Leu	Phe	Asn	
40			355					360					365				
40											GCC						1152
	Ile		Gly	Gly	Trp	Val	Ala	Ala	Gln	Leu	Ala	Pro	Pro	Ser	Ala	Ala	
		370					375					380					
4-											GCG						1200
45	Ser	Ala	Phe	Val	Gly	Ala	Gly	Ile	Ala	Gly	Ala	Ala	Val	Gly	Ser	Ile	
		•															

	385					390					395					400	
	GGC	CTC	GGG	AAG	GTG		GTG	GAC	ልሞሞ	CTG		CCT	መልመ	CCA	CCX		1248
		Leu															1240
5	2		1	-1-	405	ou		p		410	niu	GLY	TAT	GIY	415	GTÅ	
	GTG	GCA	GGC	GCG		GTG	GCC	արդ	AAG		እጥር	AGC	CCT	GAC		ccc	1296
		Ala															1230
			4	420					425			001	CLY	430	Mec	FLO	
10	TCC	ACC	GAG	GAC	CTG	GTC	AAC	TTA		CCC	GCC	ATC	CTC		ССТ	ርርጥ	1344
		Thr															1311
			435	-				440					445			1	
	GCC	CTG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	ATA	CTG	CGT	CGG	CAT	GTG	1392
15		Leu															
		450					455					460					
	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG				1431
	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu				
20	465					470					475						
	SEQ ID																
	SEQUEN						_	rs									
25	SEQUEN					acio	i									•	
	STRAND				Le												
	TOPOLOG			ar													
	ORIGIN			2													
30	ORGANI		_			, i 2011											
	IMMEDIA		-														
	CLONE:					5001	CL.										
				_													
35	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCA	GTG	ATC	GAC	TGC	AAC	ACA	TGT	GTC	48
		Tyr															
	1				5		_			10	-				15		
	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	ССТ	ACT	TTC	ACC	ATC	GAG	ACG	96
40	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
				20					25					30			
	ACG	ACC	GTA	CCC	CAA	GAT	GCG	GTG	TCG	CGC	TCG	CAG	CGG	CGA	GGC	AGG	144
	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
45			35					40					45				

	ACT	GGT	AGG	GGC	AGG	GGG	GGC	ATA	TAC	AGG	TTT	GTA	ACT	CCA	GGG	GAA	192
	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
		50					55					60					
5						TTC											240
	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Сув	Glu	Cys	Tyr	Asp	
	65					70					75					80	
	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACG	CCC	GCC	GAG	ACC	TCG	GTT	AGG	288
10	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Ser	Val	Arg	
					85					90					95		
						AAT											336
	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Сув	Gln	Asp	His	
15				100					105					110			
						AGC											384
	Leu	Glu	Phe	Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
			115					120					125				
20						ACC											432
	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	
		130					135					140					
						ACA											480
25	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Lys	Ala	Pro	Pro	Pro	
	145					150					155					160	
						TGG											528
	Ser	Trp	Asp	Gln	Met	Trp	Lys	Cys	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	
30					165					170					175		
						CTG											576
	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	
				180					185					190			
35						CCC											624
33	Val	Thr		Thr	His	Pro	Ile	Thr	Lys	Phe	Ile	Met	Ala	Сув	Met	Ser	
			195					200					205				
						GTC											672
	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	Val	Gly	Gly	Val	
40		210					215					220					
						GCG											720
		Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	Gly	Ser	Val	Val	Ile	
	225					230					235					240	
45	GTG	GGC	AGG	ATC	ATC	TTG	TCC	GGG	AGG	CCG	GCC	GTT	TTA	CCC	GAC	AGG	768

	Val	Gly	Arg	Ile	Ile 245		Ser	Gly	Arg	Pro 250		Val	Ile	Pro	Asp 255	Arg	
	GAA	GTT	CTC	TAC	CAA	GAG	TTC	GAT	GAA	ATG	GAA	GAG	TGC	GCC		CAC	816
5			Leu														
				260					265				-	270			
	CTC	CCT	TAC	ATC	GAA	CAA	GGA	ATG	CAG	CTC	GCC	GAG	CAA	TTC	AAG	CAG	864
			Tyr														
10			275					280					285		_		
	AAG	GCG	CTC	GGT	TTG	CTG	CAA	ACA	GCC	ACC	AAG	CAA	GCG	GAG	GCT	GCT	912
	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	Gln	Ala	Glu	Ala	Ala	
		290					295					300					
15	GCT	CCC	GTG	GTG	ĠAG	TCC	AAG	TGG	CGA	GCC	CTT	GAG	ACC	TTC	TGG	GCG	960
	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	Glu	Thr	Phe	Trp	Ala	
	305					310					315					320	
			ATG														1008
20	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	Tyr	Leu	Ala	Gly	Leu	
					325					330					335		
			CTG														1056
	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	Ala	Ser	Leu	Met	Ala	Phe	Thr	
25				340					345					350			
			ATC														1104
	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	Gln	Tyr	Thr	Leu	Leu	Phe	Asn	
			355					360					365				
30			GGG														1152
	Ile		Gly	Gly	Trp	Val		Ala	Gln	Leu	Ala	Pro	Pro	Ser	Ala	Ala	
		370					375					380					
			TTC														1200
35		Ala	Phe	vaı	GIĀ		Gly	Ile	Ala	Gly		Ala	Val	Gly	Ser	Ile	
	385	ome.	000			390					395					400	
			GGG														1248
	GLY	ren	Gly	гÀг		Leu	vai	Asp	He		Ala	Gly	Tyr	Gly		Gly	
40	CIEC	002	000	000	405					410					415		
			GGC														1296
	val	WIG	Gly		ren	val	АТА	Pve		Val	Met	Ser	Gly		Met	Pro	
	TICC	N.C.C	CNC	420	ome	ome			425					430			
45			GAG														1344
4 0	ser	ınr	Glu	Asp	ren	Val	Asn	Leu	Leu	Pro	Ala	Ile	Leu	Ser	Pro	Gly	

		435				440					445				
		GTC G													1392
		u Val Va	al Gly	Val		Сув	Ala	Ala	Ile	Leu	Arg	Arg	His	Val	
5	45				455					460					
		A GGG G													1431
		Gly G	lu Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu				
	465			470					475						
10															
	SEQ ID NO														
	SEQUENCE 1					:8									
	SEQUENCE ?	LAbe: un	cleic	acid											
15	STRANDEDNI	SSS: dou	ble												
	TOPOLOGY:	linear													
	Anti-sensi	E: No													
	ORIGINAL S	OURCE													
20	ORGANISM:	Hepatit	is C v	rirus											
	IMMEDIATE	EXPERIM	ENTAL	SOUR	CE										
	CLONE: N16	N15-1									•				
25	GGC TAT	ACC GG	C GAC	TTC (GAC	TCA	GTG	ATC	GAC	TGC	AAC	ACA	TGT	GTC	48
	Gly Tyr	Thr Gl	qaA y	Phe 2	qaA	Ser	Val	Ile	qaA	Cys	Asn	Thr	Cys	Val	
	1		5					10					15		
	ACC CAA	ACA GI	C GAT	TTC 2	AGC	TTG	GAC	CCT	ACT	TTC	ACC	ATC	GAG	ACG	96
20	Thr Glr	Thr Va	l Asp	Phe S	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	
30		_	0				25					30			
	ACG ACC	GTA CC	C CAA	GAT (GCG	GTG	TCG	CGC	TCG	CAG	CGG	CGA	GGC	AGG	144
	Thr Thr	Val Pr	o Gln	Asp A	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	
		35				40				•	45				
35	ACT GGT	' AGG GG	C AGG	GGG (GGC .	ATA	TAC	AGG	TTT	GTA	ACT	CCA	GGG	GAA	192
	Thr Gly	Arg Gl	y Arg	Gly (Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	
	50)			55					60					
	CGG CCC	TCA GG	C ATG	TTC (GAT '	TCT	TCG	GTC	CTG	T GT	GAA	TGT	TAT	GAC	240
40	Arg Pro	Ser Gl	y Met	Phe 1	Asp :	Ser	Ser	Val	Leu	Cys	Glu	Сув	Tyr	Asp	
	65			70					75					80	
	GCG GGC	TGT GC	T TGG	TAC (GAG (CTC .	ACG	CCC	GCC	GAG	ACC	TCG	GTT	AGG	288
	Ala Gly	Cys Al	a Trp	Tyr (Glu :	Leu	Thr	Pro	Ala	Glu	Thr	Ser	Val	Arg	
45			85					90					95		

	TTG	CGG	GCT	TAC	CTA	AAT	ACA	CCT	GGG	CTG	ccc	GTC	TGC	CAG	GAC	CAT	336
				Tyr													
				100					105					110			
5				TGG													384
	Leu	Glu	Phe	Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	
			115					120					125				
				TCC													432
10	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	
		130					135					140					
				CAG													480
		Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Lys	Ala	Pro	Pro	Pro	
15	145					150					155					160	
				CAG													528
	Ser	Trp	Asp	Gln	Met	Trp	Lys	Сув	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	
					165					170					175		
20				ACG													576
	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	
				180					185					190			
				ACA													624
25	vaı	Thr		Thr	His	Pro	Ile		Lys	Phe	Ile	Met	Ala	Cys	Met	Ser	
	com	a.a	195					200					205				
				GAG													672
	ATG		ren	Glu	vaı	Val		Ser	Thr	Trp	Val		Val	Gly	Gly	Val	
30	CMC	210	CCM	OMC.	000	999	215					220					
				CTG													720
	225	ATG	Ата	Leu	ATG		туг	Суѕ	Leu	Thr		Gly	Ser	Val	Val		
		GGC	»CC	እመሮ	አመረ	230	maa	ccc	100		235					240	
35				ATC													768
		GLY	ALY	Ile	245	пец	ser	GTÄ	Arg		Ala	Val	ITE	Pro		Arg	
34	GAA	Cum	ርሞር	ሞልሮ	_	CAC	mm/c	Cam	C22	250		a. a			255		
				TAC													816
40	014	141	Deu	Tyr 260	GIII	GIU	Pne	Авр		met	GIU	GIU	Сув		Ser	His	
	СТС	ሮሮሞ	ጥልሮ		CAA	מאם	CCA	a mor	265	OMO.	000	~~	~~~	270			
				ATC													864
			275	Ile	JLU	GIII	ату	мет 280	GTU	ren	ATG			rne	гля	GIN	
45	AAG	GCG	•	CCT	<u> </u>	СПС	CD 7		cca	N/C/C			285	a. c	-		
	-210	JCG		GGT	114	CIG	CAA	AUA	GCC	ACC	AAG	CAA	GCG	GAG	GCT	GCT	912

	Lys		Leu	Gly	Leu	Leu		Thr	Ala	Thr	Lys		Ala	Glu	Ala	Ala	
	aam.	290	ama	ama			295					300	-				
5						TCC											960
•	305	PIO	vai	vai	GIU	Ser	rĀs	Trp	Arg	Ala		Glu	Thr	Phe	Trp		
		CNC	3 mc	maa		310					315					320	
						TTC											1008
10	гÃя	urs	met	Trp		Phe	TTE	ser	GIŢ		Gln	Tyr	Leu	Ala	_	Leu	
70	mac	» Cm	cmc	CCM	325		222			330					335		
						AAC											1056
	Ser	THE	Leu		GIÀ	Asn	Pro	ALA		Ala	Ser	Leu	Met		Phe	Thr	
45	ccc	mam	N/II/CI	340	300		~~~		345					350			
15						CCG											1104
	VIG	per	355	THE	Ser	Pro	Leu		Thr	GIN	Tyr	Thr		Leu	Phe	Asn	
	እጥሮ	መመር		CCA	mcc	CmC	000	360	~~~	000			365				
						GTG											1152
20	110	370	GLY	GLY	пр	Val	375	AId	GIN	ren	Ala		Pro	ser	ALA	ATA	
	TCA		ריידירי	GTG	GGC	GCC		מיזימ	CCT	ccc	ccc	380	cmm.	cca	3.00	3.003	1200
						Ala											1200
	385				01,	390	CLI	110	niu	GIY	395	Ala	AGI	GIĀ	Ser	400	
25 .		CTC	GGG	AAG	GTG	CTT	GTG	GAC	A ጥጥ	CTG		ርርጥ	ጥልጥ	CCA	GC A		1248
						Leu											1240
	-		-	-	405					410			-1-	GIJ	415	GLY	
	GTG	GCA	GGC	GCG	CTC	GTG	GCC	TTT	AAG	_	ATG	AGC	GGT	GAC		CCC	1296
30						Val											1170
			_	420					425				 1	430			
	TCC	ACC	GAG	GAC	CTG	GTC	AAC	TTA	CTC	CCC	GCC	ATC	СТС		ССТ	GGT	1344
						Val											
35			435					440					445				
	GCC	CTG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	ATA	CTG	CGT	CGG	CAT	GTG	1392
						Val											
		450					455	•				460					
40	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG				1431
						Ala								-			
	465					470			_		475	-					

SEQ ID NO:61

55

SEQUENCE LENGTH: 2304 base pairs SEQUENCE TYPE: nucleic acid STRANDEDNESS: double

5 TOPOLOGY: linear ANTI-SENSE: NO ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

10 IMMEDIATE EXPERIMENTAL SOURCE

CLONE: N23N15A-1

	CTG	CTG	TCG	ccc	GGG	CCC	ATC	TCT	TAC	TTG	AAG	GGT	TCC	TCG	GGT	GGT	48
15	Leu	Leu	Ser	Pro	Gly	Pro	Ile	Ser	Tyr	Leu	Lys	Gly.	Ser	Ser	Gly	Gly	
	1				5					10					15		
	CCG	CTG	CCT	TGC	CCC	TCG	GGC	CGT	GTT	GTG	GGC	ATC	TTC	CGG	GCT	GCC	96
	Pro	Leu	Pro	Сув	Pro	Ser	Gly	Arg	Val	Val	Gly	Ile	Phe	Arg	Ala	Ala	
20				20					25					30			
	GTG	TGC	ACC	CGG	GGG	GTT	GCG	AAG	GCG	GTG	GAC	TTT	GTG	CCC	GTT	GAG	144
	Val	Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Val	Pro	Val	Glu	
			35					40					45				
25	TCT	ATG	GAA	ACC	ACC	ATG	CGG	TCT	CCG	GTC	TTC	ACG	GAT	AAC	TCA	ACC	192
	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	Thr	
		50					55					60					
							ACA										240
30		Pro	Ala	Val	Pro		Thr	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	
	65					70					75					80	
							ACC										288
-	Thr	Gly	Ser	Gly		Ser	Thr	Arg	Val	Pro	Ala	Ala	Tyr	Ala	Ala	Gln	
35					85					90					95		
							CTG										336
	Gly	Tyr	Lys		Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	
				100					105					110			
40							AAG										384
	Phe	Gly		Tyr	Met	Ser	Lys		His	Gly	Val	Asp	Pro	Asn	Ile	Arg	
			115					120					125				
							ACC										432
45	Thr		Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro		Thr	Tyr	Ser	Thr	
		130					135					140					

50

	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	GGT	GCC	TAT	GAC	480
				Phe													
	145					150					155				_	160	
5	ATC	ATA	ATA	TGT	GAT	GAG	TGT	CAT	TCA	ACT	GAC	TCG	ACT	TCC	ATC	TTG	528
				Cys													
					165					170					175		
	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG	GAG	ACG	GCT	GGA	GCG	CGC	CTT	576
10	Gly	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	Leu	
				180					185					190			
	GTC	GTG	CTC	GCC	ACC	GCT	ACG	CCT	CCG	GGA	TCG	GTC	ACC	GTG	CCG	CAT	624
	Val	Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	Thr	Val	Pro	His	
15			195					200					205				
	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA	GAG	ATC	CCC	TTC	672
	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Asn	Thr	Gly	Glu	Ile	Pro	Phe	
		210					215					220					
20				GCC													720
		Gly	Lys	Ala	Ile	Pro	Leu	Glu	Ala	Ile	Lys	Gly	Gly	Arg	His	Leu	
	225					230					235					240	
				CAT													768
25	Ile	Phe	Cys	His		Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	
					245					250					255		
				GGA													816
	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	qaA	Val	
30	maa			260					265					270			
				CCG													864
	ser	TTE		Pro	Thr	Ser	Gly		Val	Val	Val	Val	Ala	Thr	Asp	Ala	
	CO.	3.00	275					280					285				
35				GGC													912
	neu	290	Thr	Gly	туг	Thr		Asp	Phe	Asp	Ser		Ile	Asp	Сув	Asn	
	እሮአ		CMC	200	a. .		295					300					
				ACC													960
40	305	cys	vai	Thr	GIN		Vai	Asp	Phe	Ser		Asp	Pro	Thr	Phe	Thr	
_		CNC	3.00	100		310	~~~	~			315					320	
				ACG													1008
	116	JIU	THE	Thr		AST	Pro	GIN	Asp		Val	Ser	Arg	Ser		Arg	
45	ሮርአ	eec	NCC.	አራሙ	325	100	000	100	000	330					335		
70	CGA	GGC.	nuu	ACT	GGT'	AL÷L÷	GGC	AGG	GGG	GGC	ATA	TAC	AGG	TTT	GTA	ACT	1056

	Arg	Gly	Arg		Gly	Arg	Gly	Arg		Gly	Ile	Tyr	Arg		Val	Thr	
	CCA	000	C > 3	340	000	mas			345					350			
5			GAA														1104
3	PIO	GIY	Glu 355	Arg	PIO	ser	СТА		rne	Asp	Ser	Ser		Leu	Cys	Glu	
	ra/rm	m x m		ccc	ccc	mcm.	~~~	360	m> a				365				
			GAC														1152
40	Cys	370	Asp	ATG	GTĀ	Cys		Trp	туг	GIU	Leu		Pro	Ala	Glu	Thr	
10	mcc.		»cc	mm/c	000	COM	375	Om >				380					
			AGG														1200
	385	Val	Arg	ьеи	Arg		туг	ьеп	Asn	Thr		Gly	Leu	Pro	Val	_	
		CAC	Cam	ama	a.a	390	maa	~~~			395					400	
15			CAT														1248
	GIII	жър	His	Leu		Pne	Trp	GIU	ser		Phe	Thr	Gly	Leu		His	
	አመአ	Cam	ccc	a a a	405	mma	maa	a. a		410					415		
			GCC														1296
20	116	мвр	Ala		Pne	теп	Ser	Gin		Lys	GIn	Ala	Gly		Asn	Phe	
	ccc	ma c	CITIC	420	~~	ma a	22.5	~~~	425					430			
			CTG														1344
	PIO	туг	Leu	vaı	Ala	Tyr	GIn		Thr	Val	Cys	Ala		Ala	Lys	Ala	
25	222		435					440					445				
			CCA														1392
	Pro		Pro	Ser	Trp	Asp		Met	Trp	Lys	Сув	Leu	Ile	Arg	Leu	Lys	
		450					455					460					
30			CTA														1440
		Thr	Leu	His	Gly		Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	
	465					470					475					480	
			GAG														1488
35	GIN	Asn	Glu	vaı		Leu	Thr	His	Pro		Thr	Lys	Phe	Ile	Met	Ala	
	maa				485					490					495		
			TCG														1536
	Cys	Met	Ser		Asp	Leu	Glu	Val		Thr	Ser	Thr	Trp	Val	Leu	Val	
40				500					505					510			
70			GTC														1584
	GIĀ	GLY	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	Gly	Ser	
			515					520					525				
			ATT														1632
45	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arq	Pro	Ala	Val	Ile	

		530					535					540					
	CCC	GAC	AGG	GAA	GTT	CTC		CAA	GAG	TTC	GAT	GAA	ልጥር	GAA	GAG	ማርር	1680
												Glu					1000
5	545	-	-			550	•				555	014	1100	014	Giu	560	
·	GCC	TCG	CAC	CTC	CCT	TAC	ATC	GAA	CAA	GGA		CAG	CTC	GCC	GAG		1728
												Gln					1720
	٠				565	-				570					575	· · · ·	
10	TTC	AAG	CAG	AAG	GCG	CTC	GGT	TTG	CTG	CAA	ACA	GCC	ACC	AAG		GCG	1776
												Ala					
				580					585				•	590			
	GAG	GCT	GCT	GCT	CCC	GTG	GTG	GAG	TCC	AAG	TGG	CGA	GCC	CTT	GAG	ACC	1824
15	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	Glu	Thr	
			595					600					605				
												GGG					1872
	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	Tyr	Leu	
20		610					615					620		•			
												ATA					1920
		Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	Ala	Ser	Leu	Met	
	625					630					635					640	
25												ACC					1968
	Ala	Phe	Thr	Ala		Ile	Thr	Ser	Pro		Thr	Thr	Gln	Tyr	Thr	Leu	
	ama	mmm			645					650					655		
												CAA					2016
30	nea	File	ASII	660	ren	GIĀ	GIĀ	Trp	665	Ala	Ala	Gln	Leu		Pro	Pro	
	ACT	GCC	ር ር ጥ		GCC	መመረ	CITC	ccc		CCM	2012	GCT	000	670	00m	-	2251
												Ala					2064
			675			1110	val	680	AIG	GIY	116	ATA	685	ATG	AIG	vaı	
35	GGC	AGC		GGC	СТС	GGG	AAG		Стт	GTG	GAC	ATT		ccc	CCT	መልጥ	2112
												Ile					2112
	-	690		•		2	695				p	700	204		CLY	-1-	•
	GGA	GCA	GGG	GTG	GCA	GGC	GCG	СТС	GTG	GCC	ттт	AAG	GTC	ATG	AGC	GGT	2160
40												Lys					2200
	705		_			710					715	4 -	3_			720	
	GAC	ATG	ccc	TCC	ACC	GAG	GAC	CTG	GTC	AAC		CTC	CCC	GCC	ATC		2208
												Leu					
45					725					730					735		

	TCT	CCT	GGT	GCC	CTG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	ATA	CTG	CGT	2256
												Ala					
				740					745		_			750		•	
5	CGG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TG=	ATG	AAC	CGG	CTG	2304
												Trp					
			755					760				•	765				
	•																
10	SEQ ID	NO:	62														
	SEQUEN	CE L	ENGTI	H: 2	304	base	pai:	rs									
	SEQUEN	CE T	YPE:	nuc	leic	aci	d										
	STRAND	EDNE	ss: d	doub	le												
15	TOPOLO	GY:	linea	ar													
	ANTI-S	ense	: No														
	ORIGIN	AL S	OURCI	E													
	ORGANI	SM: I	Hepat	titi	3 C 1	viru:	s										
20	IMMEDIA	ATE 1	EXPE	RIME	TAL	sou	RCE										
	CLONE:	N231	N15B-	-1													
	CTG	CTG	TCG	CCC	GGG	CCC	ATC	TCT	TAC	TTG	AAG	GGT	TCC	TCG	GGT	GGT	48
25												Gly-					
	1				5					10	_	-			15	-	
	CCG	CTG	CCT	TGC	ccc	TCG	GGC	CGT	GTT	GTG	GGC	ATC	TTC	CGG	GCT	GCC	96
												Ilæ					
30				20					25					30			
30	GTG	TGC	ACC	CGG	GGG	GTT	GCG	AAG	GCG	GTG	GAC	TTT	GTG	CCC	GTT	GAG	144
	Val	Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phœ	Val	Pro	Val	Glu	
			35					40					45				
05	TCT	ATG	GAA	ACC	ACC	ATG	CGG	TCT	CCG	GTC	TTC	ACG	GAT	AAC	TCA	ACC	192
35												Thr					
		50					55			•		60	_				
	CCC	CCG	GCC	GTA	CCG	CAG	ACA	TTC	CAA	GTG	GCC	CAC	CTA	CAC	GCT	CCC	240
	Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	
40	65					70					75					80	
	ACT	GGC	AGC	GGC	AAA	AGC	ACC	AGG	GTG	CCG	GCT	GCG-	TAT	GCG	GCC	CAA	288
												Ala					
					85			,		9.0			•		95		
45	GGG	TAC	AAG	GTA	CTC	GTC	CTG	AAC	CCG	TCC	GTT	GCT ²	GCC	ACT		GGC	336
									_								

	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro		Val	Ala	Ala			Gly	
	անական	GGG	ccc		ልጥር	ምርር	AAC	CCA			GTT	a.a	aam	110			224
5											Val						384
			115					120					125			_	
											CCC						432
10	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro	Ile	Thr	Tyr	Ser	Thr	
		130					135					140					
	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	GGT	GCC	TAT	GAC	480
	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	
	145					150					155					160	
15	ATC	ATA	ATA	TGT	GAT	GAG	TGT	CAT	TCA	ACT	GAC	TCG	ACT	TCC	ATC	TTG	528
	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	Asp	Ser	Thr	Ser	Ile	Leu	
					165					170					175		
	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG	GAG	ACG	GCT	GGA	GCG	CGC	CTT	576
20	Gly	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	Leu	
				180					185					190			
	GTC	GTG	CTC	GCC	ACC	GCT	ACG	CCT	CCG	GGA	TCG	GTC	ACC	GTG	CCG	CAT	624
											Ser						
25			195					200					205				
	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA	GAG	ATC	CCC	TTC	672
											Thr						
		210					215					220					
30	TAT	GGC	AAG	GCC	ATC	ccc	CTC	GAG	GCC	ATC	AAG	GGG	GGG	AGG	CAT	CTC	720
,											Lys						
	225					230					235					240	
	ATT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	GCG	AAG	CTG	768
35	Ile	Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	
					245					250					255		
	TCG	GCC	CTC	GGA	GTC	AAC	GCT	GTA	GCA	TAT	TAC	CGG	GGT	CTT	GAT	GTG	816
	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	Val	
40				260					265				_	270	_		
	TCC	ATC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACA	GAC	GCT	864
											Val						
			275					280					285		•	-	
45	CTA	ATG	ACG	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCA	GTG	ATC	GAC	TGC	AAC	912
											Ser						
							-	-		•				-			

		290					295					300					
	ACA	TGT	GTC	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	CCT	ACT	TTC	ACC	960
	Thr	Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	
5	305					310					315					320	
	ATC	GAG	ACG	ACG	ACC	GTA	CCC	CAA	GAT	GCG	GTG	TCG	CGC	TCG	CAG	CGG	1008
	Ile	Glu	Thr	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	
	•				325					330					335		
10					GGT												1056
	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	Thr	
				340					345				•	350			
					CCC												1104
15	Pro	Gly		Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	Glu	
			355					360					365			•	
					GGC												1152
	Cys		Asp	Ala	Gly	Сув		Trp	Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	
20		370					375					380					
					CGG												1200
		val	Arg	ren	Arg		Tyr	Leu	Asn	Thr		Gly	Leu	Pro	Val		
	385	CAC	Cam	CMC	~~	390	maa				395					400	
25					GAG												1248
	GIII	лэр	urs	nea	Glu 405	Pne	тгр	GIU	ser	vai 410	Pne	Thr	GIĀ	ren		His	
	АТА	САТ	GCC	CAC	TTC	ጥጥረቷ	ሞርር	CAG	NCC.	_	CAC	CCX	CCN	CRC	415	mma	1206
					Phe												1296
30		-		420				O Z II	425	7170	GIII	AIG	GIY	430	VOII	FILE	
	CCC	TAC	CTG	_	GCA	TAC	CAG	GCT		GTG	ጥርሮ	CCC	ACC		MG	CCT	1344
					Ala												1311
		_	435			•		440					445		-10		
35	CCA	CCT	CCA	TCG	TGG	GAT	CAG	ATG	TGG	AAG	TGT	CTC		CGG	CTG	AAG	1392
					Trp												
		450					455		_	_	-	460		•		•	
	CCT	ACG	CTA	CAC	GGG	CCA	ACG	CCC	CTG	TTG	TAT	AGG	TTA	GGA	GCC	GTT	1440
40					Gly												
	465					470					475	-		-		480	
	CAG	AAC	GAG	GTT	ACC	CTC	ACA	CAC	CCC	ATA	ACC	AAG	TTC	ATC	ATG	GCA	1488
	Gln	Asn	Glu	Val	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	Phe	Ile	Met	Ala	
4 5					485					490					495		

	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	TGG	GTG	CTG	GTA	1536
	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	Val	
				500					505					510			
5		GGG															1584
	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Сув	Leu	Thr	Thx	Gly	Ser	
			515					520					525				
		GTC															1632
10	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	Val	Ile	
		530					535					540					
		GAC															1680
	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	Cys	
15	545					550					555					560	
		TCG															1728
	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu	Gln	
					565					570					575		
20		AAG															1776
	Phe	Lys	Gln		Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	Gln	Ala	
	a. a			580					585					590			
		GCT															1824
25	GIU	Ala		Ala	Pro	Val	Val		Ser	Lys	Trp	Arg	Ala	Leu	Glu	Thr	
	mmo.	maa	595					600					605				
		TGG															1872
	rile	Trp 610	Ala	гАг	HIS	Met		Asn	Phe	Ile	Ser		Ile	Gln	Tyr	Leu	
30	CCN		mm⁄	mcc).	áma	615			-i		620					
		GGC															1920
	625	Gly	шеш	Jei	1111	630	FLO	GIĀ	ASIL	Pro		TIE	Ala	ser	Leu		
		TTC	ACA	ccc	ጥሮጥ		NCC.	NCC.	ccc	CMC	635	3.00	<i>~</i>	ma m	200	640	1050
35		Phe															1968
					645	110		Der	110	650	1111	1111	GIII	TYL	_	ren	
	CTG	TTT	AAC	ATC		GGG	CCA	ጥርር	GTG		ccc	CD D	CTC	ccc	655	ccc	2016
		Phe															2016
40				660		U-3	01,	**P	665	A.C	niu	GIII	Leu	670	FIO	PIO	
	AGT	GCC	GCT		GCC	TTC	GTG	GGC		ርርጥ	ፈጥ ል	GCT	ccc		ር ር	ርጥጥ	2064
		Ala															2004
			675					680		1		u	685		-144	40T	
45	GGC	AGC		GGC	CTC	GGG	AAG		Chh	GTG	GAC	Aጥጥ		GCG	උඋጥ	ጥልጥ	2112
													J_0			1511	

	Gly		Ile	Gly	Leu	Gly		Val	Leu	Val	ĄsĄ		Leu	Ala	Gly	Tyr	
	CCA	690	ccc	CMC	COX	~~~	695	CMC.	Cmc	~~~		700	-			~~~	
															AGC		2160
5	705	AIG	GIY	vai	AId	710	AIG	Leu	vaı	Ald		ьys	vai	M C	Ser		
		N ITHC	ccc	mcc.	N.C.C		CNC	cmc	CBC	220	715	CmC	000	000	ATC	720	2200
															Ile		2208
10	rap	Mec	FIO	Ser	725	GIU	vab	Deu	Val	730	ъеп	rea.	PIO	ATG	735	nen	
10	ጥርጥ	ССТ	GGT	GCC		GTC	GTC	ccc	GTC		ጥረር	GCA	GCA	ልጥል	CTG	CCT	2256
															Leu		2230
			-1	740		•		021	745	,,,,	C10		,,,,,,	750	200	y	
15	CGG	CAT	GTG	GGC	CCA	GGG	GAG	GGG		GTG	CAG	TGG	ATG	. – -	CGG	CTG	2304
15															Arg		
	_		755	_		-		760				•	765		•		
20	SEQ ID	NO:	63														
20	SEQUEN	CE L	ENGTI	H: 39	564 l	oase	pair	cs									
	SEQUEN	CE T	YPE:	nuc	leic	acio	i										
	STRAND	EDNE	SS: d	duoi	le												
25	TOPOLO	GY:	linea	ar													
20	ANTI-S	ense	: No														
	ORIGIN	AL S	OURC	E								•					
	ORGANI	SM: I	Hepat	titis	S C 1	/irus	3										
30	IMMEDIA	ATE 1	EXPE	RIME	LAT	SOU	RCE										
30	CLONE:	MX2	5N15-	-1		-											
															GCC		48
35		ALA	Trp	Leu	_	Met	Met	Leu	Leu		Ala	GIn	Ala	Glu	Ala	Ala	
	1 mmc	CAC		OMO.	5	0m0	ama			10		. = 0			15		
															GCG	-	96
	Deu	GIU	ASII	20	vaı	val	ren	ABN		ATA	ser	Met	Ala		Ala	HIS	
40	ccc	ልጥሮ	CTC		mmc	cee	CIDC	anne.	25 mmc	mam	ccc	ccc	maa	30	ATC		144
															Ile		144
	GIY	116	35	Ser	FILE	rea	vaı	40	rne	Сув	Ald	Ala	45	туг	116	гуя	
	GCC	AGG		מיזירי	ር ር	GGG	ace		ጥአሮ	ር ር	יישים	መልመ	_	C/m2	TGG	ccc	192
45															Trp		136
	GLY	-ary	₽€U	AGT	110	GTÅ	AIG	vrq	TYL	vra	LIIG	TÄT	атХ	val	TTP	FLO	

		50					55					60					
	CTG	CTC	CTG	CTC	TTG	ATG	GCG	СТА	CCC	GCA	CGG	GCG	TAC	GCC	ATG	GAC	240
											Arg						
5	65					70					75					80	
	CGG	GAG	ATG	GCT	GCA	TCG	TGC	GGA	GGC	GCG	GTT	TTT	GTA	GGT	CTG	GTA	288
	Arg	Glu	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	Leu	Val	
	•				85					90					95		
10											TTC						336
	Leu	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Val	Phe	Leu	Ala	Lys	Leu	Ile	
				100					105					110			
											GAG						384
15	Trp	Trp		Gln	Tyr	Leu	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	Gln	Val	
			115					120					125				
											CGC						432
	Trp		Pro	Pro	Leu	Asn		Arg	Gly	Gly	Arg		Ala	Ile	Ile	Leu	
20		130					135					140					
											TTT						480
		Thr	Cys	ATA	Val		Pro	GIu	Leu	Ile	Phe	Asp	Ile	Thr	Lys	Leu	
	145	CITIC	caa	3.003	Om a	150	~~~				155			_		160	
25											CTC						528
	Leu	Den	ита	TTG	165	GIĂ	PIO	ren	Met		Leu	Gin	Ala	GIY		Thr	
	CAA	ልጥ/2	CCG	ጥልሮ		CMC	CCT	CCM	C2 2	170	CTC	3 mm	cam	200	175		
											Leu						576
30			110	180	11.0	V 41	мy	ATG	185	GIY	Leu	116	Arg	190	сув	Met	
	TTG	GTG	CGG		GCC	GCT	GGG	CCT		ጥልጥ	GTC	CAG	አጥር		CMC	አመረግ	624
											Val						024
			195	-			2	200		-1-			205		LCu	MCC	
35	AAG	CTG	GCT	GCA	CTG	ACA	GGT	ACG	TAC	GTT	TAT	GAC		СТТ	АСТ	CCA	672
											Tyr						0.2
		210					215		_		•	220					
	CTG	CAG	GAC	TGG	GCC	CAC	GCG	GGC	CTA	CGA	GAC	CTT	GCG	GTA	GCA	GTT	720
40											Asp						
	225					230				_	235					240	
	GAG	CCC	GTT	GCC	TTC	TCT	GAT	ATG	GAG	ACT	AAG	ATC	ATC	ACC	TGG	GGG	768
											Lys						
45					245					250					255	_	•

	GCA	GAC	ACT	GCG	GCG	TGT	GGG	GAC	ATC	ATT	TTG	GGC	CTA	CCT	GTC	TCC	816
	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro	Val	Ser	
				260					265					270			
5	GCC	CGG	AGG	GGC	AAC	GAG	ATA	CTC	CTC	GGA	CCG	GCC	GAT	AGT	TTT	GAC	864
	Ala	Arg	Arg	Gly	Asn	Glu	Ile	Leu	Leu	Gly	Pro	Ala	Asp	Ser	Phe	Asp	
			275					280					285				
	GGG	CAG	GGG	TGG	CGA	CTC	CTT	GCG	CCT	ATC	ACG	GCC	TAC	TCC	CAG	CAG	912
10	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ser	Gln	Gln	
		290					295					300					
					CTT												960
	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly	Arg	Asp	
15	305					310					315					320	
					GAG												1008
	Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	Thr	Ala	Thr	Gln	
					325					330					335		
20					ACC												1056
	Ser	Phe	Leu		Thr	Cys	Val	Asn		Val	Сув	Trp	Thr	Val	Phe	His	
				340					345					350			
					AAG												1104
25	Gly	Ala		Ser	Lys	Thr	Leu		Gly	Pro	Lys	Gly		Ile	Thr	Gln	
			355					360					365				
					GTA												1152
	Met		Thr	Asn	Val	Asp		Asp	Leu	Val	Gly	_	Ser	Ala	Pro	Pro	
30	000	370	0.00				375					380					
•					TTG												1200
	385	AId	Arg	ser	Leu	_	Pro	сув	Thr	Сув		Ser	Ser	Asp	Leu		
		CTC	NCC.	NCC.	CAM	390	CAM	CMC	3 mm	000	395	a. a				400	
35					CAT His												1248
	204	741	1111	ALY	405	ATG	мэр	Val	116	410	Val	UIR	Arg	Arg	_	Asp	
	AGC	AGG.	CCC	ACC	CTC	כיזיכ	ጥሮር	ccc	ccc		» mc	mcm.	mag	mmc	415	ccm	1206
					Leu												1296
40	502	9	Oly	420	nea	пец	261	FIO	425	PLO	TTG	Ser	TAT	430	гув	GīĀ	
	TCC	ጥርር	CCT		CCG	CTIC	CCT	mcc.		maa	ccc	CCM	cmm	-	ccc	3.00	1244
					Pro												1344
			435	321		Jeu	210	440	LIU	Ser	GTÅ	чгд	445	Val	GTÅ	TTE	
45	TTC	CGG		GCC	GTG	ጥርር	ACC		GGG	ርጥጥ	GCG	מממ		CTC	GAC	U runur	1392
						-30		-00		Gil	300	MU	GCG	GIG	anc	111	1374

	Phe	Arg 450	Ala	Ala	Val	Сув	Thr 455	Arg	Gly	Val	Ala	Lys 460	Ala	Val	Asp	Phe	
	GTG	CCC	GTT	GAG	тст	ATG	GAA	ACC	ACC	ATG	CGG	тст	CCG	GTC	TTC	ACG	1440
5															Phe		
	465					470					475					480	
	GAT	AAC	TCA	ACC	ccc	CCG	GCC	GTA	CCG	CAG	ACA	TTC	CAA	GTG	GCC	CAC	1488
	Asp	Asn	Ser	Thr	Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val	Ala	His	
10	_				485					490					495		
	CTA	CAC	GCT	CCC	ACT	GGC	AGC	GGC	AAA	AGC	ACC	AGG	GTG	CCG	GCT	GCG	1536
	Leu	His	Ala	Pro	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Arg	Val	Pro	Ala	Ala	
				500					505					510			
15	TAT	GCG	GCC	CAA	GGG	TAC	AAG	GTA	CTC	GTC	CTG	AAC	CCG	TCC	GTT	GCT	1584
	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	
			515					520					525				
	GCC	ACT	TTG	GGC	TTT	GGG	GCG	TAC	ATG	TCC	AAG	GCA	CAT	GGT	GTT	GAC	1632
20	Ala	Thr	Leu	Gly	Phe	Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Val	qaA	
		530					535					540					
	CCT	AAC	ATC	AGA	ACT	GGG	GTG	AGG	ACC	ATC	ACC	ACG	GGC	GCT	CCC	ATC	1680
	Pro	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro	Ile	
25	545					550					555					560	
	ACG	TAC	TCC	ACC	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	1728
	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	двр	Gly	Gly	Сув	Ser	Gly	
					565					570					575		
30		_			_	_				_				_	GAC		1776
	Gly	Ala	Tyr	_	Ile	Ile	Ile	Cys	_	Glu	Сув	His	Ser		Asp	Ser	
				580					585					590			
															ACG		1824
35	rnr	ser		Leu	GIĀ	TTE	GIY		Val	ren	Авр	GIN		GIU	Thr	Ala	
	CCA	cac	595	Omm.	oma	cma	O.M.O.	600		~~m		0.0m	605	CC3	maa	cmc	1072
															TCG		1872
	стХ		Arg	ьеи	vaı	Val		AIA	THE	ALG	THE		PIO	GIY	Ser	AGT	
40	N.C.C	610	ccc	Cam	ccm	220	615	CAC	CNC	CITIC	ccc	620	maa	220	a com	CCA	1020
															ACT		1920
	625	AGT	LTO	uTB	LTO	630	TIG	GIU	GIU	AGT	635	пеп	Set	noil	Thr	640	
		እጥሮ	ccc	mm/C	መለመ		አአሮ	ccc	አመሮ	ccc		CAC	ccc	አጥረግ	አአሮ		1968
45	_														AAG	_	1700
-	GIU	116	PEO	rne	TYT	стА	пÃg	WIG	116	PLO	ьeu	GTA	WIG	тте	Lys	GTĀ	

					645					650					655		
	GGG	AGG	CAT	CTC	ATT	TTC	TGC	CAT	TCC		AAG	AAA	ጥርጥ	GAC			2016
		Arg															2010
5				660			-		665		-2-	-1-	-1-	670	O_Lu	200	
	GCT	GCG	AAG	CTG	TCG	GCC	CTC	GGA	GTC	AAC	GCT	GTA	GCA	_	TAC	CGG	2064
		Ala															
			675					680					685	•	- 4		
10	GGT	CTT	GAT	GTG	TCC	ATC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	2112
		Leu															_
		690					695					700					
	GCA	ACA	GAC	GCT	CTA	ATG	ACG	GGC	TAT	ACC	GGT	GAC	TTT	GAC	TCG	GTG	2160
15		Thr															
	705					710					715			_		720	
	ATC	GAC	TGC	AAC	ACA	TGT	GTC	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	2208
	Ile	Asp	Cys	Asn	Thr	Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Авр	
20					725					730					735	_	
	CCT	ACT	TTC	ACC	ATC	GAG	ACG	ACG	ACC	GTA	ccc	CAA	GAT	GCG	GTG	TCG	2256
	Pro	Thr	Phe	Thr	Ile	Glu	Thr	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	
				740					745					750			
25		TCG															2304
	Arg	Ser	Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	
			755					760					765				
		TTT															2352
30	Arg	Phe	Val	Thr	Pro	Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	
		770					775					780					
		CTG															2400
		Leu	Cys	Glu	Сув		qaA	Ala	Gly	Сув	Ala	Trp	Tyr	Glu	Leu	Thr	
35	785					790					795					800	
		GCC															2448
	Pro	Ala	Glu	Thr		Val	Arg	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	
					805					810			,		815		
40		CCC															2496
	Leu	Pro	Val		Gln	Asp	His	Leu		Phe	Trp	Glu	Ser	Val	Phe	Thr	
	000	-		820					825					830			
		CTC															2544
4E	GLY	Leu		His	Ile	Asp	Ala		Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	
45			835					840					845				

							CTG										2592
	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	
5		850					855					860					
	AGG	GCC	AAG	GCT	CCA	CCT	CCA	TCG	TGG	GAT	CAG	ATG	TGG	AAG	TGT	CTC	2640
	Arg	Ala	Lys	Ala	Pro	Pro	Pro	Ser	Trp	Asp	Gln	Met	Trp	Lys	Cys	Leu	
	865					870					875					880	
10	ATA	CGG	CTG	AAG	CCT	ACG	CTA	CAC	GGG	CCA	ACG	CCC	CTG	TTG	TAT	AGG	2688
	Ile	Arg	Leu	Lys	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	
					885					890					895		
	TTA	GGA	GCC	GTT	CAG	AAC	GAG	GTT	ACC	CTC	ACA	CAC	CCC	ATA	ACC	AAG	2736
15	Leu	Gly	Ala	Val	Gln	Asn	Glu	${\tt Val}$	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	
				900					905					910			
	TTC	ATC	ATG	GCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	2784
	Phe	Ile	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	
20			915					920					925				
20							GTC										2832
	Trp	Val	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	
		930					935					940					
0E							ATT										2880
25		Thr	Gly	Ser	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	${ t Gly}$	Arg	
	945					950					955					960	
							AGG										2928
	Pro	Ala	Val	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	
30					965					970				•	975		
							CAC										2976
	Met	Glu	Glu	Cys	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	
				980					985					990			
35							CAG										3024
	Leu	Ala		Gln	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	
			995					.000					.005				
							GCT										3072
40			Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	
		010					.015				_	020					
	GCC																3120
	Ala	Leu	Glu	Thr	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	
45	1025					030					.035					040	
	ATA	CAG	TAC	TTA	GCA	GGC	TTG	TCC	ACT	CTG	CCT	GGA	AAC	CCC	GCA	ATA	3168

	116	GIn	Tyr	Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	
					1045					1050					1055		
	GCA	TCA	CTG	ATG	GCA	TTC	ACA	GCC	TCT	ATC	ACC	AGC	CCG	CTC	ACC	ACC	3216
5															Thr		
				1060					1065					1070			
	CAA	TAT	ACC	CTC	CTG	TTT	AAC	ATC	TTG	GGG	GGA	TGG	GTG	GCC	GCC	CAA	3264
															Ala		
10			1075					1080		_	-		1085				
	CTC	GCC	CCC	CCC	AGT	GCC	GCT	TCA	GCC	TTC	GTG	GGC	GCC	GGT	ATA	GCT	3312
															Ile		
		1090					1095					1100		-			
15	GGC	GCG	GCT	GTT	GGC	AGC	ATA	GGC	CTC	GGG	AAG	GTG	CTT	GTG	GAC	ATT	3360
															Авр		
•	1105					1110					1115					1120	
	CTG	GCG	GGT	TAT	GGA	GCA	GGG	GTG	GCA	GGC	GCG	CTC	GTG	GCC	TTT	AAG	3408
20															Phe		
					1125					L130					1135	•	
	GTC	ATG	AGC	GGT	GAC	ATG	ccc	TCC	ACC	GAG	GAC	CTG	GTC	AAC	TTA	CTC	3456
															Leu		
25				140					1145		-			1150			
	CCC	GCC	ATC	CTC	TCT	CCT	GGT	GCC	CTG	GTC	GTC	GGG			TGC	GCA	3504
															Cys		
			155					160					165		•		
30	GCA	ATA	CTG	CGT	CGG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	3552
30															Gln		
		170					.175			_		180				-	
	ATG	AAC	CGG	CTG													3564
05	Met	Asn	Arg	Leu													
35	1185																
	SEQ ID	NO: 6	4														
	SEQUENC	E LE	NGTH	: 81	.8 ba	se p	aire	}									
40	SEQUENC	E TY	PE:	nucl	.eic	acid	l										
	STRANDE	DNES	s: d	oubl	.e												
	TOPOLOG	Y: 1	inea	r									•				
	ANTI-SE	NSE:	No														
4 5	ORIGINA	L SO	URCE	}													

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE
CLONE: N22-1, N22-3, H22-8, H22-9

			,	****	٠,		o, n	22-3									
5																	
	GG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG	47
		His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu	
	•	1				5					10					15	
10	ATA	GCG	TTY	GCY	TCG	CGG	GGY	AAC	CAY	GTC	TCC	ccc	ACG	CAY	TAT	GTG	95
	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val	Ser	Pro	Thr	His	Tyr	Val	
					20					25				-	30		
	CCT	GAR	AGC	GAC	GCC	GCR	GCG	CGY	GTC	ACC	CAG	ATC	CTC	TCC	ARC	CTY	143
15	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	Xaa	Leu	
				35					40					45			
	ACC	ATC	ACT	CAG	YTG	YTG	AAG	AGG	CTY	CAC	CAG	TGG	ATT	RAT	GAK	GAC	191
	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	Gln	Trp	Ile	Asx	Xac	Asp	
20			50					55					60				
						TCY											239
	Cys	Ser	Thr	Pro	Сув	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	Asp	Trp	
		65					70					75					
25						RST											287
		Cys	Thr	Val	Leu	Xad	Asp	Xae	Lys	Thr	Trp	Leu	Gln	Ser	Lys	Leu	
	80					85					90					95	
						GGR											335
30	Leu	Pro	Arg	Leu		Gly	Val	Pro	Phe	Xaf	Ser	Сув	Gln	Arg	Gly	Tyr	
					100					105					110		
						GGA											383
	Lys	Gly	Val		Arg	Gly	Asp	Gly	Ile	Met	Xag	Thr	Thr.	Сув	Pro	Сув	
35				115					120					125			
						GGA											431
	GIĀ	Ala		Ile	Xah	Gly	His		Lys	Asn	Gly	Ser	Met	Arg	Ile	Xai	
			130					135					140				
40						AGC											479
70	Gly		Arg	Thr	Сув	Ser		Thr	Trp	Xak	Gly	Thr	Phe	Pro	Ile	Asn	
	000	145					150					155					
						CCC											527
<i>1</i> 5	ALA	тут	rnr	Thr	GLY	Pro	Cys	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	

	ARG GCG TTR TGG CGC CTR CGV RVT GAG GAG TAR GTG GAG	
	ARG GCG TTR TGG CGG GTR GCY RYT GAG GAG TAT GTG GAG GTC ACG CGG 575	5
	Xal Ala Leu Trp Arg Val Ala Xam Glu Glu Tyr Val Glu Val Thr Arg	
5	190	
	GTG GGG GAY TTC CAC TAC GTG ACG GGC ATG ACC ACT GAC AAC KTR AAA 623	ļ
	Val Gly Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Xan Lys	
	. 205	
10	TGC CCA TGC CAG GTY CCG GCC CCC GAA TTY TTC ACR GAR TTG GAT GGG 671	-
	Cys Pro Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Leu Asp Gly 210 215 220	
	COTP CCC COTP CPC ACP MAG CCM CCC CCC CCC CCC	
		ł
15	Val Arg Leu Xak Arg Tyr Ala Pro Ala Cys Lys Pro Leu Leu Arg Asp 225 230 235	
	CAC CTC ACA TIMO CAC CTC CCC CTC AAC CAA	
	Glu Val Thr Phe Gln Val Gly Leu Asn Gln Xao Xap Val Gly Ser Gln	
	240 245 250	
20	CTM CCA MCV CAC CCC CAA CCC CAA CCC	
	Leu Pro Cys Glu Pro Glu Pro Asp Val Xaq Val Val Thr Ser Met Leu	
	260 265 270	
	ACC.	
25	Thr	
25		
	Y: CorT R: AorG M: AorC K: GorT	
	S: Gor C W: A or T D: Gor Tor A	
30	Xaa : Asn or Ser	
50	Xad : Ala or Ser Xae : Cys or Phe Xaf : Phe or Leu	
	Xag : Tyr or Gln or His Xah : Thr or Ala Xai : Val or Thr	
	Xaj : Pro or Leu Xak : His or Arg Xal : Arg or Lys	
35	Xam : Ile or Ala Xan : Val or Leu Xao : Tyr or Phe	
33	Xap : Thr or Pro Xaq : Thr or Met or Ala	
	SEQ ID NO:65	
40	SEQUENCE LENGTH: 311 base pairs	
	SEQUENCE TYPE: nucleic acid	

50

45

STRANDEDNESS: double TOPOLOGY: linear ANTI-SENSE: No

ORIGINAL SOURCE

	ORGANI	SM:	Hepa	titi	s C '	viru	s										
	IMMEDI	ATE 1	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	N17	-1, 1	N17-	2, N	17-3	, H1	7-1,	H17	-3							
5																	
	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	STC	ACY	TCC	ATG	CTC	ACC	GAC	48
								Thr									
	1				5					10					15	•	
10	CCC	TCC	CAC	ATY	ACA	GCA	GAG	RCG	GCT	RRG	CGT	AGG	CTG	RCC	AGA	GGG	96
								Xab									
				20					25			_		30		-	
	TCT	CCY	CCT	YCY	TYG	RCC	AGY	TCT	TCA	GCT	AGY	CAG	TTG	TCT	GCG	СҮН	144
15								Ser									
			35					40					45				
	TCY	YYG	MAG	GCR	ACA	TGY	ACT	ACC	CAT	CAD	GRC	KCC	CCR	GAC	RCT	GAC	192
	Ser	Xae	Xag	Ala	Thr	Сув	Thr	Thr	His	Xah	Xai	Xaj	Pro	Asp	Xab	Asp	
20		50					55					60					
	CTC	ATC	GAG	GCC	AAC	CTC	CTR	TGG	CGG	CAG	GAG	ATG	GGM	GGR	AAC	ATC	240
	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	
	65					70					75					80	
25	ACC	CGY	GTG	GAG	TYA	GAG	ARC	AAG	RTA	GTR	ATT	CTR	GAC	TCT	TYY	GAM	288
	Thr	Arg	Val	Glu	Xae	Glu	Xak	Lys	Xal	Val	Ile	Leu	Asp	Ser	Xam	Xan	
					85					90					95		
	CCG	CTT	CGA	GCG	GAG	GAG	GAT	G A									311
30	Pro	Leu	Arg	Ala	Glu	Glu	Asp										
				100													
	Y :	C OI	T		R:	A or	G		M	: A	or (2		K	: G	or T	
35	S:	G or	c C		н:	A or	To	or C	D	: G	or 7	or	A				
		: Va						ıb:	Ala	or 1	hr						
		: Ar				Gly	Xε	id:	Pro	or S	Ser			Xae	: Se	er or	Leu
		: Pr					Xa	ıg:	Gln	or I	'nγs			Xah	: G1	ln or	His
40		: G1					Xa	ıj :	Ala	or S	Ser			Xak	: As	n or	Ser
	Xal	: 11	e or	· Val	Ļ		Xa	ım :	Phe	or S	er			Xan	: G1	u or	qaA
	000																
	SEQ ID						_										
45	SEQUEN	E LE	NGTE	i: 74	o ba	se r	airs	3									

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

,

	TOPOLO	GY:	line	ar													
	ANTI-S	ense	: No														
5	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa	titi	в С	viru	s										
	IMMEDI	ATE :	EXPE	RIME	LATN	SOU	RCE										
	CLONE:	028	-1, (028-	2, 0	28-4											
10																	
	GTG	GTA	GTC	CTG	GAC	TCG	TTG	GAS	CCG	CTT	CRA	GCG	RAG	GAA	GRT	GAG	48
	Val	Val	Val	Leu	qaA	Ser	Leu	Xaa	Pro	Leu	Xab	Ala	Xac	Glu	Xad	Glu	
	1				5					10					15		
15	AGG	GAA	GTG	TCC	GTT	GCG	GCG	GAG	ATC	CTG	CGR	AAG	ACC	ARG	AAA	TTC	96
	Arg	Glu	Val	Ser	Val	Ala	Ala	Glu	Ile	Leu	Arg	Lys	Thr	Xae	Lys	Phe	
				20					25					30			
	CCC	GCA	GCG	ATG	CCC	GTA	TGG	GCA	CGC	CCG	GAC	TAC	AAC	CCA	CCA	TTA	144
20	Pro	Ala	Ala	Met	Pro	Val	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	
			35					40					45				
	CTA	GAG	TCT	TGG	AAG	AAC	CCG	GAC	TAC	GTC	CCT	CCR	GTG	GTA	CAC	GGG	192
	Leu	Glu	Ser	Trp	Lys	Asn	Pro	Asp	Tyr	Val	Pro	Pro	Val	Val	His	Gly	
25		50					55					60				•	
			TTG														240
		Pro	Leu	Pro	Pro	Xaf	Lys	Ala	Pro	Pro	Ile	Pro	Pro	Pro	Arg	Arg	
	65					70					75	-				80	
30			ACG														288
	Lys	Arg	Thr	Val		Leu	Thr	Glu	Ser	Xah	Val	Ser	Ser	Ala	Leu	Ala	
		_			85					90					95		
			GCT														336
35	GLu	Leu	Ala		Lys	Thr	Phe	Gly		Ser	Gly	Ser	Ser		Val	Asp	
		-		100					105					110			
			ACG														384
	ser	GIY	Thr	АТА	Thr	Gly	Pro		Asp	Gln	Ala	Ser		Glu	Gly	Asp	
40	003	003	115	22.0				120					125				
			TCC														432
	AIG		Ser	АБР	ATA	GIU		Tyr	ser	Ser	Met		Pro	Leu	Glu	Gly	
	CNC	130	000	G 2.6	001	a>=	135					140					
45			GGG														480
70	GIU	LLO	Gly	Asp	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser	

50

	145	150	155	<u> </u>	160
	GAG GAG GCC RGC GAG			_	· -
	Glu Glu Ala Xai Glu A				
5	165	•	170		175
	ACA GGC GCC TTA ATT A	ACA CCA TGC	RCC GCG GAG		- · -
	Thr Gly Ala Leu Ile 1				
	180		185	190	
10	ATT AAT GCG CTG AGC A	AC YCT TTG	CTG CGY CAC	CAC AAC ATG	GTC TAT 624
	Ile Asn Ala Leu Ser A				
	195	200		205	-
	GCC ACA ACA TCC CGC A	GC GCA AGC	CAG CGG CAG	AAA AAG GTC	ACA TTT 672
15	Ala Thr Thr Ser Arg S	er Ala Ser	Gln Arg Gln	Lys Lys Val 1	Thr Phe
	210	215		220	
	GAC AGA CTG CAA GTC C				
	Asp Arg Leu Gln Val I	eu Asp Asp	His Tyr Arg	Asp Val Leu I	Lys Asp
20		30	235	i	240
	ATG AAG GCC AAG GCG T				740
	Met Lys Ala Lys Ala S	er			
	245	_			
25		or G	S:Gor		
	Xaa : Glu or Asp Xad : Gly or Asp		Gln or Arg	-	
	Xag : Val or Ala		Arg or Lys	_	
	Xaj : Ala or Thr		Ser or Thr Pro or Ser	Xai : Ser c	or Gly
30		AGK .	ero or ser		
	SEQ ID NO:67				
	SEQUENCE LENGTH: 515 bas	e pairs			
	SEQUENCE TYPE: nucleic a				
35	STRANDEDNESS: double				
	TOPOLOGY: linear				
	ANTI-SENSE: No				
	ORIGINAL SOURCE				
40	ORGANISM: Hepatitis C vi	rus			
	IMMEDIATE EXPERIMENTAL S	OURCE			
	CLONE: N29-1, N29-2, N29	-3			
45	AC TAC CGG GAC GTG CT	G AAG GAG A	rg aag gcg	AAG GCG TCC AC	A GTT 47

	•	Tyr .	Arg	Asp `	Val	Leu	Lys	Glu	Met	Lys	Ala	Lys	Ala	Ser	Thr '	Val	
		1				5			-		10			٠		15	
	AAG	GCT	AAA	CTT	CTA	TCT	GTA	GAG	GAA	GCC	TGY	AAG	CTG	ACG	CCC	CCA	95
5	Lys	Ala	Lys	Leu	Leu	Ser	Val	Glu	Glu	Ala	Сув	Lys	Leu	Thr	Pro	Pro	
					20					25					30		
	CAC	TCG	GCC	AGA	TCT	AAR	TTT	GGC	TAC	GGG	GCA	AAG	GAC	GTC	CGG	AGC	143
	His	Ser	Ala	Arg	Ser	Lys	Phe	Gly	Tyr	Gly	Ala	Lys	Asp	Val	Arg	Ser	
10				35					40					45			
	CTG	TCC	AGC	AAG	GCC	GTT	AAC	CAC	ATC	CGC	TCC	GTG	TGG	ARG	GAC	TTG	191
	Leu	Ser	Ser	Lys	Ala	Val	Asn	His	Ile	Arg	Ser	Val	Trp	Xaa	Asp	Leu	
			50					55					60				
15	CTG	GAA	GAC	ACT	GAR	ACA	CCA	ATT	GAC	ACC	ACC	ATC	ATG	GCA	AAA	ААТ	239
	Leu	Glu	Asp	Thr	Glu	Thr	Pro	Ile	Asp	Thr	Thr	Ile	Met	Ala	Lys	Asn	
		65					70					75					
	GAG	GTT	TTC	TGT	GTT	CAA	CCA	GAG	AAA	GGA	GGC	CGC	AAG	CCA	GCT	CGC	287
20	Glu	Val	Phe	Cys	Val	Gln	Pro	Glu	Lys	Gly	Gly	Arg	Lys	Pro	Ala	Arg	
	80					85			•		90					95	
	CTT	ATC	GTA	TTC	CCA	GAC	TTG	GGG	GTT	CGT	GTG	TGC	GAG	AAA	ATG	GCC	335
	Leu	Ile	Val	Phe	Pro	Asp	Leu	Gly	Val	Arg	Val	Cys	Glu	Lys	Met	Ala	
25					100					105					110		
	CTC	TAC	GAC	GTG	GTC	TCC	ACT	CTT	CCT	CAG	GCC	GTG	ATG	GGC	TCC	TCA	383
	Leu	Tyr	Asp	Val	Val	Ser	Thr	Leu	Pro	Gln	Ala	Val	Met	Gly	Ser	Ser	
				115					120					125			
30	TAC	GGA	TTC	CAG	TAC	TCC	CCT	GGA	CAG	CGG	GTC	GAG	TTC	CTG	GTG	AAT	431
	Tyr	Gly	Phe	Gln	Tyr	Ser	Pro	Gly	Gln	Arg	Val	Glu	Phe	Leu	Val	Asn	
			130					135					140				
	GCC	TGG	AAG	TCA	AAG	AAG	AGY	CCT	ATG	GGC	TTT	KCA	TAT	GAC	ACC	CGC	479
35	Ala	Trp	Lys	Ser	Lys	Lys	Ser	Pro	Met	Gly	Phe	Xab	Tyr	Asp	Thr	Arg	
•		145					150					155					
	TGT	TTT	GAC	TCA	ACG	GTC	ACC	GAG	AAC	GAC	ATC	CGT					515
	Cys	Phe	Asp	Ser	Thr	Val	Thr	Glu	Asn	Asp	Ile	Arg					
40	160					165					170						
40	Y:	C OI	T		R:	A or	G		K	: G	or s	r					
	Xaa	: Ly	7S 01	Glu	ı		Xá	ab:	Ala	or s	Ser						

SEQ ID NO:68

45 SEQUENCE LENGTH: 401 base pairs

55

	SEQUE	NCE 1	LABE	nuc	terc	acı	ιd										
	STRANI	DEDNI	ess:	doub	le												
	TOPOLO	GY:	line	ear													
5	ANTI-S	ENSE	: No)													
	ORIGIN	IAL S	OURC	CE													
	ORGANI	SM:	Hepa	ntiti	s C	viru	18										
	IMMEDI																
10	CLONE:	N18	3-2,	N18-	3, N	118-4	ł, н <u>і</u>	18-1,	н18	3-2,	н18-	-3					
	TG	GGG	ATC	CCG	TAT	GAT	ACC	CGC	TGC	փփփ	GAC	ጥCA	ACR	ርጥር	acv.	GAC	47
				Pro													4.7
15		1			-2	5		9	0,0		10	DGI	****	VOI	THI	15	
	ARY	GAY	ATC	CGT	RYT	GAG	GAG	TCA	АТУ	ጥልሃ		ጥሮፕ	ጣርጥ	GAC	dianc.		95
				Arg													93
		-		-	20					25		. 0,1	, cyc	p	30		
20	CCC	GAG	GCC	AGA	CAG	GCY	ATA	AGG	TCG	CTC	ACA	GAG	CGG	СТТ			143
				Arg													
				35				-	40					45			
	GGG	GGC	ccc	YTG	ACY	AAT	TCA	AAR	GGG	CAR	AAC	TGC	GGY			CGG	191
25				Leu													
			50					55				-	60			5	
	TGC	CGC	GYC	AGC	GGC	GTG	CTG	ACG	ACY	AGC	TGC	GGT	AAT	ACY	CTY	ACA	239
				Ser													
30		65		•			70					75					
-	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	TGT	CGA	GCT	GCR	AAG	CTC	CRG	GAC	287
	Cys	Tyr	Leu	Lys	Ala	Ser	Ala	Ala	Сув	Arg	Ala	Ala	Lys	Leu	Xae	Asp	
	80					85					90					95	
35	TGC	ACR	ATG	CTC	GTG	TGC	GGR	GAC	GAC	CTT	GTC	GTY	ATC	TGT	GAR	AGC	335
•	Сув	Thr	Met	Leu	Val	Cys	Gly	Asp	Asp	Leu	Val	Val	Ile	Cys	Glu	Ser	
					100					105					110		
				CAG													383
40	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	Xaa	Leu	Arg	Val	Phe	Thr	Glu	Ala	
₩				115					120					125			
	ATG	ACC	AGG	AAT	TCC	GCC											401
	Met	Thr	_	Asn	Ser	Ala											
45			130														
45	Y :	C o	r T		R:	A o	r G		H	: A	or s	ror	С				

Xaa : Asn or Ser Xab: Thr or Ile or Val Xac : Asp or Val or Ala Xad : Ala or Val Xae : Gln or Arg SEQ ID NO:69 5 SEQUENCE LENGTH: 1171 base pairs SEQUENCE TYPE: nucleic acid STRANDEDNESS: double TOPOLOGY: linear 10 ANTI-SENSE: No ORIGINAL SOURCE ORGANISM: Hepatitis C virus IMMEDIATE EXPERIMENTAL SOURCE CLONE: 030 15 TG GGG ATC CCG TAT GAT ACC CGC TGC TTT GAC TCA ACR GTC ACT GAG Gly Ile Pro Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu 20 10 AAT GAC ATC CGT GTY GAG GAG TCA ATT TAC CAA TGT TGT GAC TTG GCC 95 Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala 20 CCC GAG GCC AGA CAG GCC ATA AGG TCR CTC ACA GAG CGG CTT TAC ATC 25 Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile 40 GGG GGC CCC CTG ACT AAT TCA AAR GGG CAG AAC TGC GGY TAT CGC CGG 191 Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg 30 50 55 60 TGC CGC GYC AGC GGC GTG CTG ACG ACT AGC TGC GGY AAT ACC CTC ACA 239 Cys Arg Xaa Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr 70 35 TGT TAC TTG AAG GCC TCT GCA GCC TGT CGA GCT GCA AAG CTC CAG GAC Cys Tyr Leu Lys Ala Ser Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp 85 TGC ACG ATG CTT GTG TGC GGA GAC GAC CTT GTC GTT ATC TGT GAW AGC 40 Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Xab Ser 100 105 GCG GGA ACT CAG GAG GAC GCG GCG AGC CTA CGA GTC TTC ACG GAG GCT 383 Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu Arg Val Phe Thr Glu Ala 45 120 125

50

	ATG	ACT	AGG	TAC	TCT	GCC	ccc	ccc	GGG	GAC	CCG	CCC	CAA	CCA	GAA	TAC	431
												Pro					
			130					135					140				
5												GTG					
	Asp		Glu	Leu	Ile	Thr	Ser	Cys	Ser	Ser	Asn	Val	Ser	Val	Ala	His	
		145					150					155					
												CGT					527
10		Ala	Ser	Gly	Lys		Val	Tyr	Tyr	Leu	Thr	Arg	Asp	Pro	Xac	Thr	
	160					165					170					175	
												CAC					575
	Pro	гел	АТА	Arg		Ala	Trp	Glu	Thr		Arg	His	Thr	Pro	Val	Asn	
15	mcc.	mcc	CMA	cca	180	1 ma				185					190		
												ACC					623
	Ser	ırp	neu	195	ASII	116	TIE	met		Ala	Pro	Thr	Leu	-	Ala	Arg	
	ATG	ን ብሔ	СПС		ACC	CAC	mme	mm/c	200	3.00.0	C/mm	CTA		205	~ ~		
20												Leu					671
			210	1100		******	1110	215	Ser	TTG	neu	ьец	220	GIN	GIU	GIN	
	CTT	GAA		GCC	CTA	GAT	TGT	_	Aጥር	ጥልሦ	ccc	GCC	_	ሞልሮ	mee	አ መጠ	719
05												Ala					113
25		225	-			-	230			-4-	1	235		-1-	502		
	GAG	CCA	CTT	GAC	CTA	CCT	CAG	ATC	ATT	CAA	CGA	CTC	CAY	GGT	СТТ	AGC	767
												Leu					
30	240					245					250			-		255	
00	GCA	TTT	TCA	CTC	CAT	AGT	TAC	TCT	CCA	GGT	GAG	ATC	AAT	AGG	GTG	GCT	815
	Ala	Phe	Ser	Leu	His	Ser	Tyr	Ser	Pro	Gly	Glu	Ile	Asn	Arg	Val	Ala	
					260					265					270		
35												CGA					863
	Ser	Суѕ	Leu		Lys	Leu	Gly	Val	Pro	Pro	Leu	Arg	Val	Trp	Arg	His	
				275					280					285			
												CAG					911
40	Arg	Ala		Ser	Val	Arg	Ala	Lys	Leu	Leu	Ser	Gln	Gly	Gly	Arg	Ala	
	000		290					295					300	•			
												GTA					959
	WIG		cys	GIĀ	гла	Tyr		Phe	Asn	Trp	Ala	Val	Lys	Thr	Lys	Leu	
45	***	305	3.00	005			310					315					
.=	AAA	CYC	ACT	CCA	ATC	CCR	GAA	GCG	TCC	CAG	CTG	GAC	TTG	TCC	GGC	TGG	1007

	Lys Xad Thr Pro Ile Pro Glu Ala Ser Gln Leu Asp Leu Ser Gly Trp 320 325 330 335	
	TTC GTT GCT GGT TAC AGC GGG GGA GAC ATA TAT CAC AGC CTG TCT CGT 10	055
5	Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg	
	340 345 350	
	GCC CGA CCC CGC TGG TTY ATG TGG TGC CTA CTC CTA CTT TCC GTA GGG 1	103
	Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly	
10	355 360 365	
	GTA GGC ATC TAC CTG CTC CCC AAC CGA TGA GCG GGG AGC TAA ACA CTC 1	151
	Val Gly Ile Tyr Leu Leu Pro Asn Arg StopAla Gly Ser StopThr Leu 370 375 380	
15	Gln Ala Asn Arg Pro Ser	171
	385	
	Y: CorT R: AorG M: AorC W: AorT	
20	Xaa : Val or Ala Xab : Asp or Glu Xac : Thr or Pro	
20	Xad : Leu or Pro	
	SEQ ID NO:70	
	SEQUENCE LENGTH: 1084 base pairs	
25	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
	ANTI-SENSE: No	
30	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
35	CLONE: 2217	
33		
	GG CAT GTG GGC CCA GGG GAG GGG GCT GTG CAG TGG ATG AAC CGG CTG	47
	His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu	
40	1 5 10 15	
	ATA GCG TTT GCT TCG CGG GGC AAC CAT GTC TCC CCC ACG CAC TAT GTG	95
	Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val	
	20 25 30	
45		143
	Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Asn Leu	
50		

				35					40					45			
	ACC	ATC	ACT	CAG	CTG	TTG	AAG	AGG	-	CAC	CAG	TGG	ል ጥጥ		GAG	GAC	191
										His							131
5			50				-	55					60	21011	014	rwp	
	TGC	TCC	ACG	CCA	TGC	TCC	GGC	TCG	TGG	CTC	AGG	GAT		TGG	GAC	TGG	239
										Leu							
		65					70		_		•	75					
10	ATA	TGC	ACG	GTA	TTG	GCT	GAT	TTC	AAG	ACC	TGG	CTC	CAG	TCC	AAG	CTC	287
										Thr							
	80					85					90		•		_	95	
	CTG	CCG	CGG	TTA	CCG	GGG	GTC	CCT	TTT	TTC	TCA	TGC	CAG	CGT	GGG	TAC	335
15	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	Ser	Сув	Gln	Arg	Gly	Tyr	
					100					105					110		
	AAG	GGG	GTT	TGG	CGG	GGA	GAT	GGC	ATC	ATG	TAT	ACC	ACC	TGC	CCA	TGT	383
	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	Tyr	Thr	Thr	Cys	Pro	Сув	
20				115					120					125			
										AAC							431
	Gly	Ala		Ile	Thr	Gly	His	Val	Lys	Asn	Gly	Ser	Met	Arg	Ile	Val	
			130					135					140	•			
25										CAC							479
	GIĀ		Arg	Thr	Cys	Ser		Thr	Trp	His	Gly	Thr	Phe	Pro	Ile	Asn	
	COO	145					150					155					
										TCC							527
30	160	ığı	THE	THE	GIĀ		Cys	ınr	Pro	Ser		Ala	Pro	Asn	Tyr		
		CCC	መጥረ፤	WCC.	ccc	165 CTC	ccc	3 (1101)	CNC	C1 C	170	ama				175	
										GAG Glu							575
	9				180	vai	nia	116	GIU	185	TAT	Val	GIU	vai	190	Arg	
35	GTG	GGG	GAT	TTC		TAC	ርጥር	ACG	ccc	ATG	ACC	ልሮሞ	GAC	አአሮ		222	622
										Met							623
		- 4		195		-1-			200	1100	****	1112	nop	205	VAI	пуь	
	TGC	CCA	TGC		GTT	CCG	GCC	CCC		TTC	TTC	ACA	GAA		CAT	ccc	671
40										Phe							0,1
			210					215					220		тор	U.J	
	GTG	CGG	CTG	CAC	AGG	TAC	GCT		GCG	TGC	AAA	ССТ		CTG	CGG	GAT	719
										Cys							
45		225					230				-	235			-	-	

	GAG	GTC	ACA	TTC	CAG	GTC	GGG	CTC	AAC	CAA	ТАТ	ACG	GTT	GGG	тса	CAG	767
		Val															
	240					245	_				250			1	-	255	
5	CTC	CCA	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA		GTC	ACC	TCC	ATG		815
		Pro															015
					260			-		265					270		
	ACC	GAC	CCC	TCC	CAC	ATT	ACA	GCA	GAG	GCG	GCT	AGG	CGT	AGG		ACC	863
10		Asp															000
		_		275					280				9	285			
	AGA	GGG	тст	CCC	CCT	TCC	TCG	ACC	AGT	тст	TCA	GCT	AGT		ጥጥር	ጥርጥ	911
		Gly															711
15			290					295					300				
10	GCG	CTT	TCT	TCG	CAG	GCA	ACA	TGC	ACT	ACC	CAT	CAG		GCC	CCA	GAC	959
		Leu															
		305					310	_				315					
20	ACT	GAC	CTC	ATC	GAG	GCC	AAC	CTC	CTG	TGG	CGG	CAG	GAG	ATG	GGC	GGA	1007
20		Asp															
	320					325				_	330			•		335	
	AAC	ATC	ACC	CGC	GTG	GAG	TCA	GAG	AAC	AAG	ATA	GTA	ATT	CTA	GAC	TCT	1055
		Ile															
25					340					345					350		
	TTT	GAA	CCG	CTT	CGA	GCG	GAG	GAG	GAT	GA							1084
	Phe	Glu	Pro	Leu	Arg	Ala	Glu	Glu	Asp						•		
				355					360								
30																	
	SEQ ID	NO: 7	1														

SEQUENCE LENGTH: 1004 base pairs

SEQUENCE TYPE: nucleic acid

35 STRANDEDNESS: double

TOPOLOGY: linear ANTI-SENSE: No

ORIGINAL SOURCE

40 ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

CLONE: 1728

45 TGT GAG CCC GAA CCG GAT GTA ACA GTG GTC ACC TCC ATG CTC ACC GAC 48

55

	Cys 1	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val			Ser	Met	Leu		Asp	
		mee	CAC	Amm	5 202	COL	020	000		10					15		
5																GGG	96
	110	DCI	111.5	20	****	VIG	GIU	Ala	25	Arg	Arg	Arg.	Leu	Inr 30	Arg	Gly	
	тст	CCC	ССТ		TCG	ACC	АСТ	ጥርቱ		CCT	እርጥ	CAG	mm/c		ccc	C ID ID	144
	-				Ser												144
10			35					40	501	*****	OCL	0111	45	Ser	NIG	Dea	
	TCT	TCG	CAG	GCA	ACA	TGC	ACT		CAT	CAG	GGC	GCC		GAC	ACT	GAC	192
					Thr												
		50				_	55					60					
15	CTC	ATC	GAG	GCC	AAC	CTC	CTG	TGG	CGG	CAG	GAG	ATG	GGC	GGA	AAC	ATC	240
					Asn												
	65					70					75					.80	
					TCA												288
20	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	Ile	Val	Ile	Leu	Asp	Ser	Phe	Glu	
					85					90					95		
					GAG												336
	Pro	Leu	Arg		Glu	Glu	Asp	Glu		Glu	Val	Ser	Val	Ala	Ala	Glu	
25	3.00	ama		100					105					110			
					ACC												384
	116	reu	115	гля	Thr	Arg	Lys		Pro	Ala	Ala	Met		Val	Trp	Ala	
	CGC	CCG		መልሮ	AAC	CCN	CCX	120	CMB	CNC	mam	maa	125		000	a. a	420
30					Asn												432
	5	130		-1-		110	135	neu	Deu	Giu	Ser	140	тÃр	Abii	PIO	Asp	
	TAC	GTC	CCT	CCA	GTG	GTA		GGG	TGC	CCA	TTG		ССТ	ACC	AAG	GCC	480
					Val												
35	145					150		-	-		155				•	160	
	CCT	CCA	ATA	CCA	ССТ	CCA	CGA	AGA	AAG	AGA	ACG	GTT	GTC	CTG	ACA	GAA	528
					Pro												
					165					170					175		
40	TCC	TCC	GTG	TCC	TCT	GCC	TTG	GCG	GAG	CTT	GCT	ACA	AAG	ACC	TTT	GGC	576
	Ser	Ser	Val	Ser	Ser	Ala	Leu	Ala	Glu	Leu	Ala	Thr	Lys	Thr	Phe	Gly	
				180					185					190			
-					TCG												624
45	Ser	Ser	Gly	Ser	Ser	Ala	Val	Asp	Ser	Gly	Thr	Ala	Thr	Gly	Pro	Pro	•

			195					200					205				
	GAC	CAG	GCC	TCC	GCC	GAA	GGA	GAT	GCA	GGA	TCC	GAC	GCT	GAG	TCG	TAC	672
_															Ser		
5		210					215					220				_	
	TCC	TCC	ATG	CCC	CCC	CTT	GAG	GGA	GAG	CCG	GGG	GAC	CCC	GAT	CTC	AGC	720
	Ser	Ser	Met	Pro	Pro	Leu	Glu	Gly	Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	
	225					230					235		•			240	
10															GTC		768
	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser	Glu	Glu	Ala	Ser	Glu	Asp	Val	Val	
					245					250					255		
															CCA		816
15	Сув	Сув	Ser		Ser	Tyr	Thr	Trp	Thr	Gly	Ala	Leu	Ile	Thr	Pro	Сув	
				260					265					270			
															CCT		864
	Ala	Ala		Glu	Ser	Lys	Leu		Ile	Asn	Ala	Leu	Ser	Asn	Pro	Leu	
20			275					280					285				
															GCA		912
	Leu		His	His	Asn	Met		Tyr	Ala	Thr	Thr	Ser	Arg	Ser	Ala	Ser	
	~~~	290					295					300					
25															GAT		960
		Arg	GIN	гля	rys		Thr	Phe	Asp	Arg		Gln	Val	Leu	Asp		
	305	ma 0	000	~~~	-	310					315					320	
						CTC									AC		1004
30	ure	туr	Arg	двр		Leu	тÃŝ	Asp	Met		Ala	Lys	Ala	Ser			
					325					330							

SEQ ID NO:72

SEQUENCE LENGTH: 857 base pairs 35 SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: No 40

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE

CLONE: 2918 45

50

	AC			GAC													47
			Arg	Asp	Val	Leu	Lys	Asp	Met	Lys	Ala	Lys	Ala	Ser	Thr	Val	
		1				5					10					15	·
5	AAG	GCI	' AAA	CTT	CTA	TCI	GTA	GAG	GAA	GCC	TGC	AAG	CTC	AC	CCC	CCA	95
	Lys	Ala	Lys	Leu	Leu	Ser	. Val	Glu	Glu	Ala	с Сув	Lys	Lev	Thi	: Pro	Pro	
					20					25					30		
																AGC	143
10	His	Ser	Ala	Arg	Ser	Lys	Phe	Asp	Tyr	Gly	Ala	Lys	Asp	Val	Gln	Ser	
				35					40	ı				45	5		
	CTG	TCC	AGC	AAG	GCC	GTT	AAC	CAC	ATC	CAC	TCC	GTG	TGG	AAG	GAC	TTG	191
	Leu	Ser	Ser	Lys	Ala	Val	Asn	His	Ile	His	Ser	Val	Trp	Lys	Asp	Leu	
15			50					55					60	)			
	CCG	GAA	GAC	ACT	GAG	ACA	CCA	ATC	GAC	ACC	ACC	ATC	ATG	GCA	AAA	AAT	239
	Pro	Glu	_Asp	Thr	Glu	Thr	Pro	Ile	Asp	Thr	Thr	Ile	Met	Ala	Lys	Asn	
		65					70					75					
20	GAG	GTT	TTT	TGT	GTT	CAA	CCA	GAG	AAA	GGA	GGC	CGC	AAG	CCA	GCT	CGC	287
20	Glu	Val	Phe	Cys	Val	Gln	Pro	Glu	Lys	Gly	Gly	Arg	Lys	Pro	Ala	Arg	
	80					85					90					95	
	CTT	ATC	GTA	TTC	CCA	GAC	TTG	GGG	GTT	CGT	GTG	TGC	GAG	AAA	ATG	GCC	335
				Phe													
25					100					105				_	110		
	CTC	TAC	GAC	GTG	GTC	TCC	ACT	CTT	CCT	CAG	GCC	GTG	ATG	GGC	TCC	TCA	383
				Val													
				115					120					125			
30	TAC	AGA	TTT	CAG	TGC	TCC	CCT	GGA	CAG	CGG	GTC	GAG	TTC	CTG	GTG	AAT	431
				Gln													
			130					135		_			140				
	GCC	TGG	AAG	TCA	AAG	AAG	AGC	CCT	ATG	GGC	TTT	GCA	TAT	GAC	ACC	CGC	479
35				Ser													
		145					150			_		155	-	•			
۸.	TGT	TTT	GAC	TCA	ACG	GTC	ACC	GAG	AAC	GAC	ATC	CGT	ACT	GAG	GAG	TCA	527
				Ser													
40	160					165					170	•				175	
	ATT	TAT	CAA	<b>T</b> GT	TGT	GAC	TTG	GAC	CCC	GAG	GCC	AGA	CAG	GCC	ATA		575
				Cys													- · •
					180	_		_		185		,			190	J	
45	TCG	CTC	ACA	GAG	CGG	СТТ	ТАТ	ATC	GGG		CCC	TTG	ACC	ልልጥ		444	623
															·	-HILL	043

	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	Gly	Gly	Pro	Leu	Thr	Asn	Ser	Lys	
				195	_		_		200	-				205		-1-	
	GGG	CAA	AAC	TGC	GGC	TAT	CGC	CGG	TGC	CGC	GCC	AGC	GGC	GTG	CTG	ACG	671
5																Thr	• -
			210					215		_			220				
	ACT	AGC	TGC	GGT	AAT	ACC	CTC	ACA	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	719
																Ala	
10		225					230					235					
		CGA															767
	Cys	Arg	Ala	Ala	Lys	Leu	Gln	Asp	Сув	Thr	Met	Leu	Val	Cys	Gly	Asp	
	240					245					250					255	
15																GCA	815
	Asp	Leu	Val	Val		Cys	Glu	Ser	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	
					260					265					270		
		CTA															857
20	Abn	Leu	Arg		Phe	Thr	Glu	Ala		Thr	Arg	Asn	Ser				
				275					280					285			
	SEQ ID	NO: 7	13														
	SEQUEN			T: 15	R18 F	1886	nair	~a									
25	SEQUEN						_	.5									
	STRANDI					4010	_										
	TOPOLO																
	ANTI-SI	ENSE:	No														
30	ORIGINA	AL SC	URCE	3													
	ORGANIS	SM: H	lepat	itis	C v	rirus	3										
	IMMEDIA	ATE E	XPEF	RIMEN	ITAL	SOUF	RCE										
	CLONE:	1718	:														
35																	
	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	GTC	ACC	TCC	ATG	CTC	ACC	GAC	48
	Сув	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	Thr	Asp	
	1				5					10					15		
40		TCC															96
	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Arg	Arg	Arg	Leu	Thr	Arg	Gly	
				20					25					30			
		CCC															144
45	Ser	Pro	Pro	Ser	Ser	Thr	Ser	Ser	Ser	Ala	Ser	Gln	Leu	Ser	Ala	Leu	

			35		•			40					A E				
	TCT	TCG		GCA	ACA	ጥርር	ልሮሞ	_	Cam	CAC	ccc	cco	45			GAC	100
	Ser	Ser	Gln	Ala	Thr	Cvs	Thr	Thr	Uic	CAG	GGC	31.	CCA	GAC	ACT	GAC	192
5		50				Cyb	55		шта	GIII	GIY		Pro	Asp	Thr	Asp	
	СТС		GAG	GCC	AAC	כיזיכי			ccc	CNC	C2C	60	000				
			Glu														240
	65				-1011	70	Je u	11p	ALG	GIII	75	Met	GIY	GIA	Asn		
10	_	CGC	GTG	GAG	TCA		אמר	AAC	מיזיא	CIDA		CON	CNG	mam	mmm	80	
••			Val														288
					85	0		LJU	110	90	116	neu	мър	261		GIU	
	CCG	CTT	CGA	GCG		GAG	GAT	GAG	AGG		GTG	ጥሮሮ	ርምጥ	ccc	95 ccc	CNC	226
15			Arg														336
			-	100			2		105		•	001	141	110	AIG	Giu	
	ATC	CTG	CGG	AAG	ACC	AGG	AAA	TTC	-	GCA	GCG	ATG	CCC		TICC	CCA	384
			Arg														304
20			115	_		_	•	120					125	•42	*TP	AIG	
20	CGC	CCG	GAC	TAC	AAC	CCA	CCA	TTA	CTA	GAG	TCT	TGG	_	AAC	CCG	GAC	432
			Asp														432
		130					135					140	_1 -				
25	TAC	GTC	CCT	CCA	GTG	GTA	CAC	GGG	TGC	CCA	TTG		ССТ	ACC	AAG	GCC	480
23			Pro														
	145					150		_	_		155					160	
	CCT	CCA	ATA	CCA	CCT	CCA	CGA	AGA	AAG	AGA	ACG	GTT	GTC	CTG	ACA		528
30			Ile														
30					165					170			·		175		
	TCC	TCC	GTG	TCC	TCT	GCC	TTG	GCG	GAG	CTT	GCT	ACA	AAG	ACC	TTT	GGC	576
	Ser	Ser	Val	Ser	Ser	Ala	Leu	Ala	Glu	Leu	Ala	Thr	Lys	Thr	Phe	Gly	
25				180					185					190			
35			GGA														624
	Ser	Ser	Gly	Ser	Ser	Ala	Val	Asp	Ser	Gly	Thr	Ala	Thr	Gly	Pro	Pro	
			195					200					205				
			GCC														672
40	Asp		Ala	Ser	Ala	Glu	Gly	Asp	Ala	Gly	Ser	Asp	Ala	Glu	Ser	Tyr	
		210					215					220					
			ATG														720
		Ser	Met	Pro	Pro	Leu	Glu	Gly	Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	
45	225					230					235		•			240	

	GAC	GGG	TCT	TGG	тст	ACC	GTA	AGC	GAG	GAG	GCC	AGC	GAG	GAC	GTC	GTC	768
												Ser					
					245					250					255		
5	TGC	TGC	TCG	ATG	TCC	TAC	ACA	TGG	ACA	GGC	GCC	TTA	ATT	ACA	CCA	TGC	816
	Cys	Cys	Ser	Met	Ser	Tyr	Thr	Trp	Thr	Gly	Ala	Leu	Ile	Thr	Pro	Cys	
				260					265	•				270			
												CTG					864
10	Ala	Ala	Glu	Glu	Ser	Lys	Leu	Pro	Ile	Asn	Ala	Leu	Ser	Asn	Pro	Leu	
			275					280					285				
												TCC					912
	Leu		His	His	Asn	Met	Val	Tyr	Ala	Thr	Thr	Ser	Arg	Ser	Ala	Ser	
15		290					295					300					
												CAA					960
		Arg	GIn	Lys	Lys		Thr	Phe	yab	Arg	Leu	Gln	Val	Leu	Asp	Asp	
	305	m. 0				310					315					320	
20																	1008
	UIR	туг	Arg	Asp		Xaa	ràs	Asp	Met		Ala	Lys	Ala	Ser		Val	
	አልሮ	CCT	***	CMM	325	man	<i></i>	~~~		330					335		
																	1056
25	270	AIG	пуз	340	Deu	Ser	Val	GIU	345	AIA	Сув	Lys	Leu		Pro	Pro	
	CAC	TCG	GCC		тст	AAA	արարար	GAC		GGG	CCA	220	CAC	350	C a C	300	1104
												Lys					1104
			355			-1-		360	-1-	011	71.4	ny o	365	Val	GIII	Ser	
30	CTG	TCC	AGC	AAG	GCC	GTT	AAC		ATC	CAC	TCC	GTG		AAG	GAC	ጥጥር	1152
												Val					1132
		370					375					380			<b>F</b>		
	CCG	GAA	GAC	ACT	GAG	ACA	CCA	ATC	GAC	ACC	ACC	ATC	ATG	GCA	AAA	AAT	1200
35												Ile					
	385					390					395				-	400	
	GAG	GTT	TTT	TGT	GTT	CAA	CCA	GAG	AAA	GGA	GGC	CGC	AAG	CCA	GCT	CGC	1248
												Arg					
40					405					410					415		
	CTT	ATC	GTA	TTC	CCA	GAC	TTG	GGG	GTT	CGT	GTG	TGC	GAG	AAA	ATG	GCC	1296
	Leu	Ile	Val	Phe	Pro	Asp	Leu	Gly	Val	Arg	Val	Cys	Glu	Lys	Met	Ala	
				420					425					430			
45	CTC	TAC	GAC	GTG	GTC	TCC	ACT	CTT	CCT	CAG	GCC	GTG	ATG	GGC	TCC	TCA	1344

	Leu	Tyr			Val	Ser	Thr		Pro	Gln	Ala	Val	Met	Gly	Ser	Ser	
			435					440					445				
																	1392
5	Tyr	Arg	Phe	Gln	Суз	Ser	Pro	Gly	Gln	Arg	Val	Glu	Phe	Leu	Val	Asn	
		450					455					460					
																	1440
	Ala	Trp	Lys	Ser	Lys	Lys	Ser	Pro	Met	Gly	Phe	Ala	Tyr	Asp	Thr	Arg	
10	465					470					475					480	
	TGT	TTT	GAC	TCA	ACG	GTC	ACC	GAG	AAC	GAC	ATC	CGT	ACT	GAG	GAG	TCA	1488
	Cys	Phe	Asp	Ser	Thr	Val	Thr	Glu	Asn	Asp	Ile	Arg	Thr	Glu	Glu	Ser	
					485					490					495		
15	ATT	TAT	CAA	TGT	TGT	GAC	TTG	GAC	CCC	GAG	GCC	AGA	CAG	GCC	ATA	AGG	1536
		Tyr															
				500					505					510		_	
	TCG	CTC	ACA	GAG	CGG	CTT	TAT	ATC	GGG	GGC	CCC	TTG	ACC	ААТ	TCA	AAA	1584
20		Leu															
			515					520	_	_			525				
	GGG	CAA	AAC	TGC	GGC	TAT	CGC	CGG	TGC	CGC	GCC	AGC	GGC	GTG	CTG	ACG	1632
		Gln															
25		530				_	535	_	-	•		540	- 4				
20	ACT	AGC	TGC	GGT	AAT	ACC	CTC	ACA	TGT	TAC	TTG	AAG	GCC	тст	GCA	GCC	1680
		Ser															
	545					550			•	•	555					560	
	TGT	CGA	GCT	GCG	AAG	CTC	CAG	GAC	TGC	ACG		CTC	GTG	TGC	GGA		1728
30		Arg															2720
					565			•	4	570					575	·····p	
	GAC	СТТ	GTC	GTT	ATC	TGT	GAA	AGC	GCG		ACC	CAG	GAG	GAC		CCA	1776
		Leu															1770
35	-			580		-1-			585	011		Ų1II	GIU	590	AIG	AIG	
	AAC	СТА	CGA		ттс	ACG	GAG	ርር ጥ		ACC	NGC	አአጥ	ጥሮሮ				1010
		Leu															1818
		- ·· <del>-</del>	595					600	.100	TITT	AL Y	VOII	605	wig			
40								500					005				

SEQ ID NO:74

SEQUENCE LENGTH: 2591 base pairs

SEQUENCE TYPE: nucleic acid

45 STRANDEDNESS: double

55

TOPOLOGY: linear
ANTI-SENSE: NO
ORIGINAL SOURCE
ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE
CLONE: 2218

.0	GG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG	47
															Arg		
		1				5					10	~			5	15	
4.5	ATA	GCG	TTI	GCT	TCG	CGG	GGC	AAC	CAT	GTC	: TCC	ccc	ACG	CAC	TAT	GTG	95
15															Туг		
					20					25					30		
	CCI	GAA	AGC	GAC	GCC	GCA	GCG	CGC	GTC	ACC	CAG	ATC	CTC	TCC	AAC	CTT	143
•	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	Asn	Leu	
20				35					40					45			
	ACC	ATC	ACT	CAG	CTG	TTG	AAG	AGG	CTT	CAC	CAG	TGG	ATT	AAT	GAG	GAC	191
	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	Gln	Trp	Ile	Asn	Glu	Asp	,
05			50					55					60				
25															GAC		239
	Cys			Pro	Сув	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	Asp	Trp	
		65					70					75					
00															AAG		287
30			Thr	Val	Leu	_	Asp	Phe	Lys	Thr	Trp	Leu	Gl'n	Ser	Lys	Leu	
	80					85					90					95	
															GGG		335
0.5	reu	Pro	Arg	Leu		Gly	Val	Pro	Phe		Ser	Суѕ	Gln	Arg	Gly	Tyr	
35	220	ccc	C mm	maa	100		~			105					110		
															CCA		383
	пĵо	GIY	, val	11p	Arg	GIĀ	Asp	GIY		Met	Tyr	Thr	Thr		Pro	Cys	
	CCA	CCA	CAR		N.C.C	CC3	@3.m	ama	120					125			. 0
40															ATC		431
	GLY	VIG	130	116	1111	GIĀ	urs		гĀ2	ASN	GIĀ	ser		Arg	Ile	Val	
	GGG	CCT		acc.	m∕cm	NCC.	220	135	maa	020	001		140				
															ATC		479
45	~~1	145	Arg	TIIT	Cys	Ser	150	THE	TED	urs	стХ		Pne	Pro	Ile	Asn	
		*43					TOU					155					

		TAC	ACC	ACA	GGC	CCC	TGC	ACA	CCC	TCC	CCG	GCG	CCA	AAC	TAT	TCC	527
																	321
	Ala	Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	
	160					165					170					175	
5	AGG	GCG	TTG	TGG	CGG	GTG	GCC	ATT	GAG	GAG	TAT	GTG	GAG	GTC	ACG	CGG	575
	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ile	Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	
					180					185					190		
	GTG	GGG	GAT	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	GTG	AAA	623
10	Val	Gly	Asp	Phe	His	Tyr	Val	Thr	Gly	Met	Thr	Thr	Asp	Asn	Val	Lys	
				195					200				_	205			
	TGC	CCA	TGC	CAG	GTT	CCG	GCC	CCC	GAA	TTC	TTC	ACA	GAA	TTG	GAT	GGG	671
	Суз	Pro	Cys	Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	Gly	
15			210					215					220				
				CAC													719
	Val	Arg	Leu	His	Arg	Tyr	Ala	Pro	Ala	Cys	Lys	Pro	Leu	Leu	Arg	Asp	
		225					230					235					
20				TTC													767
	Glu	Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Tyr	Thr	Val	Gly	Ser	Gln	
	240					245					250					255	
				GAG													815
25	Leu	Pro	Сув	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	
					260					265					270		
				TCC													863
	Thr	Asp	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Arg	Arg	Arg	Leu	Thr	
30				275					280	*				285			
				CCC													911
	Arg	Gly		Pro	Pro	Ser	Ser		Ser	Ser	Ser	Ala	Ser	Gln	Leu	Ser	
			290					295					300				
35				TCG													959
	Ala		Ser	Ser	Gln	Ala		Сув	Thr	Thr	His	Gln	Gly	Ala	Pro	Asp	
		305			_		310					315					
																	1007
40		Asp	Leu	Ile	Glu		Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	
40	320					325					330					335	
																	1055
	Asn	шe	Thr	Arg		Glu	Ser	Glu	Asn		Ile	Val	Ile	Leu	Asp	Ser	
45					340					345					350		
<b>4</b> 5	TTT	GAA	CCG	CTT	CGA	GCG	GAG	GAG	GAT	GAG	AGG	GAA	GTG	TCC	GTT	GCG	1103

	Phe	Glu	Pro			Ala	Glu	Glu	Asp	Glu	Arg	Glu	Val	Ser	Val	Ala	
				355					360					365			
																	1151
5	Ala	Glu		Leu	Arg	Lys	Thr	Arg	Lys	Phe	Pro	Ala	Ala	Met	Pro	Val	
			370					375					380				
																	1199
	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	Leu	Glu	Ser	Trp	Lys	Asn	
10		385					390					395	•				
																	1247
	Pro	Asp	Tyr	Val	Pro	Pro	Val	Val	His	Gly	Сув	Pro	Leu	Pro	Pro	Thr	
	400					405					410					415	
15	AAG	GCC	CCT	CCA	ATA	CCA	CCT	CCA	CGA	AGA	AAG	AGA	ACG	GTT	GTC	CTG	1295
	Lys	Ala	Pro	Pro	Ile	Pro	Pro	Pro	Arg	Arg	Lys	Arg	Thr	Val	Val	Leu	
					420					425					430		
	ACA	GAA	TCC	TCC	GTG	TCC	TCT	GCC	TTG	GCG	GAG	CTT	GCT	ACA	AAG	ACC	1343
20						Ser											
20				435					440				-	445			
	TTT	GGC	agt	TCC	GGA	TCG	TCG	GCC	GTC	GAC	AGC	GGC	ACG	GCG	ACC	GGC	1391
	Phe	Gly	Ser	Ser	Gly	Ser	Ser	Ala	Val	Asp	Ser	Gly	Thr	Ala	Thr	Gly	
25			450					455					460			_	
23	CCT	CCT	GAC	CAG	GCC	TCC	GCC	GAA	GGA	GAT	GCA	GGA	TCC	GAC	GCT	GAG	1439
	Pro	Pro	Asp	Gln	Ala	Ser	Ala	Glu	Gly	Авр	Ala	Gly	Ser	Asp	Ala	Glu	
		465					470					475	•				
	TCG	TAC	TCC	TCC	ATG	CCC	CCC	CTT	GAG	GGA	GAG	CCG	GGG	GAC	ccc	GAT	1487
30						Pro											
	480					485					490					495	
	CTC	AGC	GAC	GGG	TCT	TGG	TCT	ACC	GTA	AGC	GAG	GAG	GCC	AGC	GAG	GAC	1535
						Trp											
35					500					505					510	_	
	GTC	GTC	TGC	TGC	TCG	ATG	TCC	TAC	ACA	TGG	ACA	GGC	GCC	TTA	ATT	ACA	1583
						Met											
				515					520	_		_		525			
40	CCA	TGC	GCC	GCG	GAG	GAG	AGC	AAG	CTG	CCC	ATT	AAT	GCG	CTG	AGC	AAC	1631
						Glu											
			530					535					540				
	CCT	TTG	CTG	CGC	CAC	CAC	AAC	ATG	GTC	TAT	GCC	ACA	ACA	TCC	CGC	AGC	1679
45						His											=

		545					550					555		•			
	GCA	AGC	CAG	CGG	CAG	AAA	AAG	GTC	ACA	TTT	GAC	AGA	CTG	CAA	GTC	CTG	1727
		Ser															
5	560					565					570					575	
	GAT	GAC	CAC	TAC	CGG	GAC	GTG	CTS	AAG	GAC	ATG	AAG	GCC	AAG	GCG	TCC	1775
	Asp	Asp	His	Tyr	Arg	Asp	Val	Xaa	Lys	Asp	Met	Lys	Ala	Lys	Ala	Ser	
					580					585					590		
10																	1823
	Thr	Val	Lys	Ala	Lys	Leu	Leu	Ser	Val	Glu	Glu	Ala	Cys	Lys	Leu	Thr	
				595					600					605			
																	1871
15	Pro	Pro		Ser	Ala	Arg	Ser	Lys	Phe	Asp	Tyr	Gly	Ala	Lys	Asp	Val	
			610					615					620				
																	1919
	Gln	Ser	Leu	Ser	Ser	Lys		Val	Asn	His	Ile		Ser	Val	Trp	Lys	
20	~~~	625					630					635					
																	1967
	640	Leu	PLO	GIU	Asp		GTA	Thr	Pro	TTE		Thr	Thr	Ile	Met		
		ከልጥ	GAG	രത്ത	mmm	645	cmm	C2.2	003	<b>C2</b> C	650	001	000	000		655	0015
25		Asn															2015
	2,0		0.14	Val	660	Cys	AGI	GIII	PIO	665	пув	GIY	GTĀ	Arg	670	PIO	
	GCT	CGC	CTT	ATC		TTC	CCA	GAC	ጥጥር		CTT	CGT	GTG	ጥርር		מממ	2063
		Arg															2003
30		_		675					680	2		9		685		2,0	
	ATG	GCC	CTC	TAC	GAC	GTG	GTC	TCC	ACT	СТТ	ССТ	CAG	GCC		ATG	GGC	2111
		Ala															
			690					695					700			-	
35	TCC	TCA	TAC	AGA	TTT	CAG	TGC	TCC	CCT	GGA	CAG	CGG	GTC	GAG	TTC	CTG	2159
	Ser	Ser	Tyr	Arg	Phe	Gln	Cys	Ser	Pro	Gly	Gln	Arg	Val	Glu	Phe	Leu	
		705					710					715					
	GTG	AAT	GCC	TGG	AAG	TCA	AAG	AAG	AGC	CCT	ATG.	GGC	TTT	GCA	TAT	GAC	2207
40		Asn	Ala	Trp	Lys	Ser	Lys	Lys	Ser	Pro	Met	Gly	Phe	Ala	Tyr	qaA	
	720					725					730					735	
																	2255
	Thr	Arg	Сув	Phe	Asp	Ser	Thr	Val	Thr	Glu	Asn	Asp	Ile	Arg	Thr	Glu	
45					740					745					750		

	GAG	TCA	ATT	TAT	CAA	TGT	TGT	GAC	TTG	GAC	CCC	GAG	GCC	AGA	CAG	GCC	2303
	Glu	Ser	Ile	Tyr	Gln	Cys	Cys	Asp	Leu	Asp	Pro	Glu	Ala	Arg	Gln	Ala	
				755					760					765			
5	ATA	AGG	TCG	CTC	ACA	GAG	CGG	CTT	TAT	ATC	GGG	GGC	CCC	TTG	ACC	AAT	2351
	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	Gly	Gly	Pro	Leu	Thr	Asn	
			770					775					780				
	TCA	AAA	GGG	CAA	AAC	TGC	GGC	TAT	CGC	CGG	TGC	CGC	GCC	AGC	GGC	GTG	2399
10		Lys															
		785					790					795			_		
	CTG	ACG	ACT	AGC	TGC	GGT	AAT	ACC	CTC	ACA	TGT	TAC	TTG	AAG	GCC	TCT	2447
46	Leu	Thr	Thr	Ser	Сув	Gly	Asn	Thr	Leu	Thr	Сув	Tyr	Leu	Lys	Ala	Ser	
15	800					805					810					815	·
	GCA	GCC	TGT	CGA	GCT	GCG	AAG	CTC	CAG	GAC	TGC	ACG	ATG	CTC	GTG	TGC	2495
	Ala	Ala	Cys	Arg	Ala	Ala	Lys	Leu	Gln	Asp	Сув	Thr	Met	Leu	Val	Cys	
20					820					825					830		
-	GGA	GAC	GAC	CTT	GTC	GTT	ATC	TGT	GAA	AGC	GCG	GGA	ACC	CAG	GAG	GAC	2543
	Gly	Asp	Asp	Leu	Val	Val	Ile	Cys	Glu	Ser	Ala	Gly	Thr	Gln	Glu	Asp	
				835					840					845			
25	GCG	GCA	AAC	CTA	CGA	GTC	TTC	ACG	GAG	GCT	ATG	ACC	AGG	ААТ	TCC	GCC	2591
	Ala	Ala	Asn	Leu	Arg	Val	Phe	Thr	Glu	Ala	Met	Thr	Arg	Asn	Ser	Ala	
			850					855					860				

SEQ ID NO:75

SEQUENCE LENGTH: 4296 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: No

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

O IMMEDIATE EXPERIMENTAL SOURCE

**CLONE: 1530U** 

GCGGATCCT CCA CCT CCA TCG TGG GAT CAA ATG TGG AAG TGT CTC ATA CGG 51

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg

1 5 10

CTG AAG CCT ACG CTA CAC GGG CCA ACG CCC CTG TTG TAT AGG TTA GGA 99

50

			Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	
	15					20					25					30	
																ATC	147
5	Ala	Val	Gln	Asn	Glu	Val	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	Phe	Ile	
					35					40					45		
																GTG	195
	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	
10				50					55					60			
	CTG																243
	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	
			65					70					75				
15					ATT												291
	Gly	Ser	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	
		80					85					90					
					AGG												339
20	Val	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	
	95					100					105					110	
	GAG	TGC	GCC	TCG	CAC	CTC	CCT	TAC	ATC	GAA	CAA	GGA	ATG	CAG	CTC	GCC	387
	Glu	Сув	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	
25					115					120					125		
					CAG												435
	Glu	Gln	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Lys	
				130					135					140			
30					GCT												483
50	Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	
			145					150					155				
					GCG												531
35	Glu		Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	
33		160					165					170					
					TTG												579
-		Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	Ala	Ser	
	175					180					185					190	
40					ACA												627
	Leu	Met	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	Gln	Tyr	
					195					200					205		
					AAC												675
45	Thr	Leu	Leu	Phe	Asn	Ile	Leu	Gly	Gly	Trp	Val	Ala	Ala	Gln	Leu	Ala	

				210					215					220			
	CCC	CCC	AGT	GCC	GCT	TCA	GCC	TTC	GTG	GGC	GCC	GGT	АТА		GGC	CCC	723
															Gly		
5			225					230		-		2	235		017		
	GCT	GTT	GGC	AGC	ATA	GGC	CTC	GGG	AAG	GTG	CTT	GTG		ATT	CTG	GCG	771
															Leu		
	•	240				_	245	•	. 4			250	P		204	1114	
10	GGT	TAT	GGA	GCA	GGG	GTG	GCA	GGC	GCG	CTC	GTG		ттт	AAG	GTC	ATG	819
															Val		023
	255				_	260		_			265			-1-		270	
	AGC	GGT	GAC	ATG	ccc	TCC	ACC	GAG	GAC	CTG		AAC	TTA	CTC	CCC		867
15															Pro		•
					275				_	280					285		
	ATC	CTC	TCT	CCT	GGT	GCC	CTG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	АТА	915
	Ile	Leu	Ser	Pro	Gly	Ala	Leu	Val	Val	Gly	Val	Val	Cys	Ala	Ala	Ile	
20				290					295	-			-	300			
	CTG	CGT	CGG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	963
															Met		
			305					310		_			315	•			
25	CGG	CTG	ATA	GCG	TTT	GCT	TCG	CGG	GGC	AAC	CAT	GTC	TCC	CCC	ACG	CAC	1011
-0															Thr		
		320					325					330					
	TAT	GTG	CCT	GAA	AGC	GAC	GCC	GÇA	GCG	CGC	GTC	ACC	CAG	ATC	CTC	TCC	1059
30	Tyr	Val	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	
30	335					340					345					350	
	AAC	CTT	ACC	ATC	ACT	CAG	CTG	TTG	AAG	AGG	CTT	CAC	CAG	TGG	ATT	AAT	1107
	Asn	Leu	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	Gln	Trp	Ile	Asn	
05					355					360					365		
35	GAG	GAC	TGC	TCC	ACG	CCA	TGC	TCC	GGC	TCG	TGG	CTC	AGG	GAT	GTT	TGG	1155
	Glu	Asp	Сув	Ser	Thr	Pro	Сув	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	
				370					375					380			
	GAC	TGG	ATA	TGC	ACG	GTA	TTG	GCT	GAT	TTC	AAG	ACC	TGG	CTC	CAG	TCC	1203
40	Asp	Trp	Ile	Cys	Thr	Val	Leu	Ala	Asp	Phe	Lys	Thr	Trp	Leu	Gln	Ser	
			385					390					395				
																	1251
	Lys	Leu	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	Ser	Cys	Gln	Arg	
45		400					405					410				=	

		GGG	TAC	AAG	GGG	GTT	TGG	CGG	GGA	CAT	GGC	ልጥሮ	እጥር	תאת	NCC.	N.C.C	שכיר	1299
		Glv	Tvr	Lvs	Glv	Val	Trp	Ara	Glv	Agn	Glv	Tlo	Met	TEL	mb ~	Wha	Crrc	1233
		415	_	•	4		420	9	O.J	110p	GLY	425	Mec	ıyı	1111	1111	430	
	5	CCA	TGT	GGA	GCA	CAA		ACC	GGA	САТ	GTC		ልልሮ	CCT	መርጣ	አጥር		1347
													Asn					1347
				2		435			011	113.0	440	ыyв	Non	GIY	261	445	vrd	
		ATC	GTT	GGG	ССТ		ACC	ጥርጥ	AGC	ልልሮ		ሞሮር	Cac	CCA	እሮአ		ccc	1395
1	0												His					1333
•	·				450	5		-1-	001	455	****		1110	GIY	460	rne	PLO	
		ATC	AAC	GCG	TAC	ACC	ACA	GGC	ccc		ACA	ccc	TCC	CCG		CCA	AAC	1443
													Ser					1773
1	5			465	_			_	470	-4 -				475				
•	•	TAT	TCC	AGG	GCG	TTG	TGG	CGG	GTG	GCC	ATT	GAG	GAG		GTG	GAG	GTC	1491
													Glu					
			480				_	485					490	•				
2	0	ACG	CGG	GTG	GGG	GAT	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	1539
2	U												Met					
		495					500					505				-	510	
		GTG	AAA	TGC	CCA	TGC	CAG	GTT	CCG	GCC	CCC	GAA	TTC	TTC	ACA	GAA	TTG	1587
2	<b>.</b>												Phe					
2	3					515					520					525		
																		1635
		Asp	Gly	Val	Arg	Leu	His	Arg	Tyr	Ala	Pro	Ala	Сув	Lys	Pro	Leu	Leu	
30	^				530					535					540			
30	U																	1683
		Arg	Asp		Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Tyr	Thr	Val	Gly	
				545					550					555				
-	-																	1731
3	•	Ser		Leu	Pro	Сув	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	
			560					565					570					
																		1779
	_		Leu	Thr	Asp	Pro		His	Ile	Thr	Ala	Glu	Ala	Ala	Arg	Arg	Arg	
4(	0	575					580					585					590	
																		1827
		Leu	Thr	Arg	Gly		Pro	Pro	Ser	Ser	Thr	Ser	Ser	Ser	Ala	Ser	Gln	
	_	<b></b>				595					600					605		
45	5	TTG	TCT	GCG	CTT	TCT	TCG	CAG	GCA	ACA	TGC	ACT	ACC	CAT	CAG	GGC	GCC	1875

	Leu	Ser	Ala	Leu	Ser	Ser	Gln	Ala	Thr	Cys	Thr	Thr	His	Gln	Gly	Ala	
				610					615					620			
	CCA	GAC	ACT	GAC	CTC	ATC	GAG	GCC	AAC	CTC	CTG	TGG	CGG	CAG	GAG	ATG	1923
5	Pro	Asp	Thr	Asp	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	
			625					630					635				
	GGC	GGA	AAC	ATC	ACC	CGC	GTG	GAG	TCA	GAG	AAC	AAG	ATA	GTA	ATT	CTA	1971
	Gly	Gly	Asn	Ile	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	Ile	Val	Ile	Leu	
10		640					645					650					
	GAC	TCT	TTT	GAA	CCG	CTT	CGA	GCG	GAG	GAG	GAT	GAG	AGG	GAA	GTG	TCC	2019
						Leu											
	655					660					665					670	
15	GTT	GCG	GCG	GAG	ATC	CTG	CGG	AAG	ACC	AGG	AAA	TTC	CCC	GCA	GCG	ATG	2067
	Val	Ala	Ala	Glu	Ile	Leu	Arg	Lys	Thr	Arg	Lys	Phe	Pro	Ala	Ala	Met	
					675					680					685		
	CCC	GTA	TGG	GCA	CGC	CCG	GAC	TAC	AAC	CCA	CCA	TTA	CTA	GAG	TCT	TGG	2115
20	Pro	Val	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	Leu	Glu	Ser	Trp	
_				690					695					700			
	AAG	AAC	CCG	GAC	TAC	GTC	CCT	CCA	GTG	GTA	CAC	GGG	TGC	CCA	TTG	CCG	2163
	Lys	Asn	Pro	Asp	Tyr	Val	Pro	Pro	Val	Va:1	His	Gly	Cys	Pro	Leu	Pro	
25			705					710					715				
	CCT	ACC	AAG	GCC	CCT	CCA	ATA	CCA	CCT	CCA	CGA	AGA	AAG	AGA	ACG	GTT	2211
	Pro	Thr	Lys	Ala	Pro	Pro	Ile	Pro	Pro	Pro	Arg	Arg	Lys	Arg	Thr	Val	
		720					725					730					
30	GTC	CTG	ACA	GAA	TCC	TCC	GTG	TCC	TCT	GCC	TTG	GCG	GAG	CTT	GCT	ACA	2259
30	Val	Leu	Thr	Glu	Ser	Ser	Val	Ser	Ser	Ala	Leu	Ala	Glu	Leu	Ala	Thr	
	735					740					745					750	
																	2307
05	Lys	Thr	Phe	Gly	Ser	Ser	Gly	Ser	Ser	Ala	Val	Asp	Ser	Gly	Thr	Ala	
35					755					760					765		
																	2355
	Thr	Gly	Pro	Pro	Asp	Gln	Ala	Ser	Ala	Glu	Gly	Asp	Ala	Gly	Ser	Asp	
				770					775					780			
40																	2403
	Ala	Glu	Ser	Tyr	Ser	Ser	Met	Pro	Pro	Leu	Glu	Gly	Glu	Pro	Gly	Asp	
			785					790					795				
																	2451
<b>4</b> 5	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser	Glu	Glu	Ala	Ser	

		800					805					810					
	GAG	GAC	GTC	GTC	TGC	TGC	TCG	ATG	TCC	TAC	ACA	-	ACA	GGC	GCC	ጥጥል	2499
												Trp					
5	815					820				-	825			2		830	
	ATT	ACA	CCA	TGC	GCC	GCG	GAG	GAG	AGC	AAG	CTG	ccc	ATT	AAT	GCG		2547
												Pro					
	•				835					840					845		
10	AGC	AAC	CCT	TTG	CTG	CGC	CAC	CAC	AAC	ATG	GTC	TAT	GCC	ACA	ACA	TCC	2595
												Tyr					
				850					855				•	860			
	CGC	AGC	GCA	AGC	CAG	CGG	CAG	AAA	AAG	GTC	ACA	TTT	GAC	AGA	CTG	CAA	2643
15	Arg	Ser	Ala	Ser	Gln	Arg	Gln	Lys	Lys	Val	Thr	Phe	Asp	Arg	Leu	Gln	
			865					870					875				
																	2691
	Val		Asp	Asp	His	Tyr	Arg	Asp	Val	Leu	Lys	Asp	Met	Lys	Ala	Lys	
20		880					885					890					
																	2739
		Ser	Thr	Val	Lys		Lys	Leu	Leu	Ser		Glu	Glu	Ala	Cys	Lys	
	895	200	000	~~>		900					905					910	
25																	2787
	rea	THE	Pro	Pro		ser	Ala	Arg	Ser		Phe	Asp	Tyr	Gly		Lys	
	GAC	ርሞር	CNC	NCC.	915	maa	300		000	920					925		
																	2835
30	ımp	vui	GIN	930	Den	Ser	per	пув	935	val	Asn	His	TTE		ser	vaı	
	TGG	AAG	GAC		CCG	GAA	GAC	ልሮሞ		እርን	CCA	Ame.	CNC	940	300	3 m/a	2883
												Ile					2003
	-	_	945				F	950	<b></b>		110	116	955	****	1111	116	
35	ATG	GCA	AAA	ААТ	GAG	GTT	TTT		GTT	CAA	CCA	GAG		CCA	GGC	CGC	2931
												Glu					2,51
		960					965	_				970		2	1	9	
	AAG	CCA	GCT	CGC	CTT	ATC	GTA	TTC	CCA	GAC	TTG		GTT	CGT	GTG	TGC	2979
40												Gly					
	975					980				-	985	•		,		990	
	GAG	AAA	ATG	GCC	CTC	TAC	GAC	GTG	GTC	TCC	ACT	CTT	CCT	CAG	GCC		3027
												Leu					
<b>4</b> 5					995					000		*			005		

	ATG	GGC	TCC	TCA	TAC	AGA	$\mathbf{T}\mathbf{T}\mathbf{T}$	CAG	TGC	TCC	CCT	GGA	CAG	CGG	GTC	GAG	3075
																Glu	
				1010					1015					1020			
5	TTC	CTG	GTG	AAT	GCC	TGG	AAG	TCA	AAG	AAG	AGC	CCT	ATG	GGC	TTT	GCA	3123
	Phe	Leu	Val	Asn	Ala	Trp	Lys	Ser	Lys	Lys	Ser	Pro	Met	Gly	Phe	Ala	
			1025					1030					1035				
																	3171
10			Thr	Arg	Сув	Phe	Asp	Ser	Thr	Val	Thr	Glu	Asn	Asp	Ile	Arg	
		1040					1045					1050					
																	3219
		Glu	Glu	Ser	Ile	Tyr	Gln	Cys	Сув	Asp	Leu	Asp	Pro	Glu	Ala	Arg	
15	1055					1060					1065					1070	
																	3267
	Gln	Ala	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	Gly	Gly	Pro	Leu	
					1075					1080					1085		
20																	3315
	Thr	Asn	Ser		Gly	Gln	Asn			Tyr	Arg	Arg	Cys	Arg	Val	Ser	
				090					95					100			
																	3363
25	GIY		Leu	Thr	Thr	Ser			Asn	Thr	Leu			Tyr	Leu	Lys	
	ccc		1105	000	mem.	001		1110					1115				
																	3411
		120	Ala	WIG	сув			Ата	ràs	Leu			Сув	Thr	Met	Leu	
30			CCA	GAC	CAC		125	C mm	3.000	mam		1130					
			Gly														3459
	1135	O, D	O _T	мър		140	Vai	Val	116		Asp 1145	ser	Ala	GIĀ			
		GAC	GCG	GCG			CGA	ርጥር	புரு			CCM	እመሮ	» CM		1150 mag	3507
35			Ala														3507
					.155	Deu	Arg	Vai		1111	Giu	мта	Met		AEG	ıyr	
	TCT	GCC	CCC			GAC	CCG	ccc			CAA	መእሮ	CAC			CMC	3555
			Pro														3333
40				170	2				.175	110	Giu	TYL		180	GIU	rea	
	ATA	ACA			TCC	TCC	ТАА			GTC	GCG	CAC			ጥሮል	CCC	3603
			Ser														3003
			185	_		-		190					195		JUL	-~J	
45	AAA	CGG	GTG	TAC	TAT	CTC			GAC	CCC	ACC			СТА	GCG	CGG	3651
												_					

	Lys	Arg	Val	Tyr	Тут	Leu	Thr	Arg	Asp	Pro	Thr	Thr	Pro	Leu	Ala	Arg	
		1200					1205					1210					
																	3699
5	Ala	Ala	Trp	Glu	Thr	Ala	Arg	His	Thr	Pro	Val	Asn	Ser	Trp	Leu	Gly	
	1215					1220					1225					1230	
																	3747
	Asn	Ile	Ile	Met	Tyr	Ala	Pro	Thr	Leu	Trp	Ala	Arg	Met	Ile	Leu	Met	
10					1235					1240		-			1245		
																	3795
	Thr	His	Phe	Phe	Ser	Ile	Leu	Leu	Ala	Gln	Glu	Gln	Leu	Glu	Lys	Ala	
				1250					1255					1260			
15																	3843
	Leu	Asp	Cys	Gln	Ile	Tyr	Gly	Ala	Thr	Tyr	Ser	Ile	Glu	Pro	Leu	Asp	
			1265					1270					1275				
																	3891
20	Leu	Pro	Gln	Ile	Ile	Gln	Arg	Leu	His	Gly	Leu	Ser	Ala	Phe	Ser	Leu	
		1280					1285					1290					
																	3939
	His	Ser	Tyr	Ser	Pro	Gly	Glu	Ile	Asn	Arg	Val	Ala	Ser	Сув	Leu	Arg	
25	1295					1300					1305					1310	
																	3987
	Lys	Leu	Gly	Val	Pro	Pro	Leu	Arg	Val	Trp	Arg	His	Arg	Ala	Arg	Ser	
					1315					1320					1325	•	
30																	4035
	Val	Arg	Ala	Lys	Leu	Leu	Ser	Gln	Gly	Gly	Arg	Ala	Ala	Thr	Cys	Gly	
				1330					L335					1340			
																	4083
35	Lys	Tyr	Leu	Phe	Asn	Trp	Ala	Val	Lys	Thr	Lys	Leu	Lys	Leu	Thr	Pro	
00			1345					1350					L35 <b>5</b>				
																	4131
			Glu	Ala	Ser	Gln	Leu	Asp	Leu	Ser	Gly	Trp	Phe	Val	Ala	Gly	
40		1360					1365					.370					
40																	4179
		Ser	Gly	Gly	Asp	Ile	Tyr	His	Ser	Leu	Ser	Arg	Ala	Arg	Pro	Arg	
	1375					1380					385					390	
																	4227
45	Trp	Phe	Met	Trp	Cys	Leu	Leu	Leu	Leu	Ser	Val	Gly	Val	Gly	Ile	Tvr	

o*				e	139	-				140		The state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the state of the s	***	£,	140		
								GGG	GAGC	TAAA	CA C	TCCA	GGCC	A AT	AGGC	CATC	4280
_	re	и ье	u Pr			g St	ор										
5	CC	CCMM	TTTT	141													
	CC	CCII	1111	111	*11												4296
	SEQ I	D NO	•76														
10	SEQUE	- 2		<b>тн</b> •	818	hago	nai	~e									
10	SEQUE						-	1.5									
	STRAN					- uc											
	TOPOL	OGY:	lin	ear													
15	ANTI-	SENS	E: N	0													
	ORIGI	NAL	SOUR	CE	•												
	ORGAN	ISM:	Нер	atit.	is C	vir	us										
	IMMED	IATE	EXP	ERIM	ENTA	L SO	URCE								•		
20	CLONE	: N2	2-1														
	GG		GTG														47
			Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu	
25		1				5				-	10					15	
			TTT														95
	TIE	ATS	Phe	Ala		Arg	Gly	Asn	His		Ser	Pro	Thr	His	_	Val	
	CCT	CAA	NCC.	CAC	20	<i>a</i> a.	000	000		25					30	_	
30			AGC														143
	110	O.L.u	Ser	35	Ald	AIG	ALG	ALG	40	THE	GIN	тте	Leu		Asn	Leu	
	ACC	ATC	ACT		CTG	ттс	AAG	AGG		CAC	CAG	ሞርር	<b>አ</b> ጥጥ	45 מחממ	CAC	CAC	191
			Thr														131
35			50				•	55					60		<b>01</b> 4	шър	
	TGC	TCC	ACG	CCA	TGC	TCC	GGC	TCG	TGG	CTC	AGG	GAT		TGG	GAC	TGG	239
			Thr														
		65					70					75		_	-	-	
40	ATA	TGC	ACG	GTA	TTG	GCT	GAT	TGC	AAG	ACC	TGG	CTC	CAG	TCC	AAG	CTC	287
	Ile	Cys	Thr	Val	Leu	Ala	Asp	Cys	Lys	Thr	Trp	Leu	Gln	Ser	Lys	Leu	
	80					85					90					95	
			CGG														335
<b>4</b> 5	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	Ser	Cys	Gln	Arg	Gly	Tyr	

					100					105					110		
	AAG	GGG	GTT	TGG	CGG	GGA	GAT	GGC	ATC	ATG	TAT	ACC	ACC	TGC	CCA	TGT	383
							Asp										
5				115					120					125		-	
	GGA	GCA	CAA	ATC	ACC	GGA	CAT	GTC	AAA	AAC	GGT	TCT	ATG	AGG	ATC	GTT	431
							His										
	•		130					135					140	_			
10	GGG	CCT	AGA	ACC	TGT	AGC	AAC	ACG	TGG	CAC	GGA	ACA	TTT	CCC	ATC	AAC	479
							Asn										
		145					150					155					
	GCG	TAC	ACC	ACA	GGC	CCC	TGC	ACA	CCC	TCC	CCG	GCG	CCA	AAC	TAT	TCC	527
15	Ala	Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	
	160					165					170		,		_	175	
	AGG	GCG	TTG	TGG	CGG	GTG	GCC	ATT	GAG	GAG	TAT	GTG	GAG	GTC	ACG	CGG	575
	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ile	Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	
20 ·					180					185					190	_	
	GTG	GGG	GAT	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	GTG	AAA	623
	Val	Gly	Asp	Phe	His	Tyr	Val	Thr	Gly	Met	Thr	Thr	qaA	Asn	Val	Lys	
				195					200					205			
25	TGC	CCA	TGC	CAG	GTT	CCG	GCC	CCC	GAA	TTC	TTC	ACA	GAA	TTG	GAT	GGG	671
	Cys	Pro	Cys	Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	Gly	
			210					215					220				
	GTG	CGG	CTG	CAC	AGG	TAC	GCT	CCG	GCG	TGC	AAA	CCT	CTC	CTG	CGG	GAT	719
30	Val	Arg	Leu	His	Arg	Tyr	Ala	Pro	Ala	Cys	Lys	Pro	Leu	Leu	Arg	Asp	
		225					230					235					
	GAG	GTC	ACA	TTC	CAG	GTC	GGG	CTC	AAC	CAA	TAT	ACG	GTT	GGG	TCA	CAG	767
	Glu	Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Tyr	Thr	Val	Gly	Ser	Gln	
35	240					245					250					255	
00	CTC	CCA	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	GTC	ACC	TCC	ATG	CTC	815
	Leu	Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	
					260					265					270		
40	ACC																818
<del></del> -0	Thr																

SEQ ID NO:77

SEQUENCE LENGTH: 818 base pairs SEQUENCE TYPE: nucleic acid

55

50

	STRANDEDNESS: double
	TOPOLOGY: linear
	ANTI-SENSE: NO
5	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
	IMMEDIATE EXPERIMENTAL SOURCE
	CLONE: N22-3
10	

	GG	CAI	GIG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG	47
		His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu	
		1				5					10					15	
15	ATA	GCG	TTT	GCT	TCG	CGG	GGC	AAC	CAT	GTC	TCC	CCC	ACG	CAC	TAT	GTG	95
	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val	Ser	Pro	Thr	His	Tyr	Val	
					20					25			•		30		
	CCT	GAA	AGC	GAC	GCC	GCA	GCG	CGC	GTC	ACC	CAG	ATC	CTC	TCC	AAC	CTT	143
20	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	Asn	Leu	
				35					40					45			
	ACC	ATC	ACT	CAG	TTG	TTG	AAG	AGG	CTC	CAC	CAG	TGG	ATT	AAT	GAG	GAC	191
	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	Gln	Trp	Ile	Asn	Glu	qsA	
25			50					55					60				
								TCG									239
	Cys		Thr	Pro	Cys	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	Asp	Trp	
		65					70					75					
30								TTC									287
••		Сув	Thr	Val	Leu	Ala	Asp	Phe	Lys	Thr	Trp	Leu	Gln	Ser	Lув	Leu	
	80					85					90					95	
								CCT									335
35	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	Ser	Cys	Gln	Arg	Gly	Tyr	
35					100					105					110		
								GGC									383
	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	Tyr	Thr	Thr	Cys	Pro	Сув	
				115					120					125			
40								GTC									431
	Gly	Ala	Gln	Ile	Thr	Gly	His	Val	Lys	Asn	Gly	Ser	Met	Arg	Ile	Val	
			130					135					140				
	GGG	CTT	AGA	ACC	TGT	AGC	AAC	ACG	TGG	CAC	GGA	ACA	TTC	CCC	ATC	AAC	479
45	Gly	Leu	Arg	Thr	Сув	Ser	Asn	Thr	Trp	His	Gly	Thr	Phe	Pro	Ile	Asn	

		145					150					155					
	GCG	TAC	ACC	ACA	GGC	ccc	TGC	ACA	ccc	TCT	CCA	GCG	CCG	AAC	TAC	TCC	527
																Ser	
5	160					165					170					175	
											TAT						575
	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ala	Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	
					180					185					190		
10																AAA	623
	Val	Gly	Asp		His	Tyr	Val	Thr	Gly	Met	Thr	Thr	Авр	Asn	Val	Lys	
				195					200					205			
											TTC						671
15	Сув	Pro		GIn	Val	Pro	Ala	_	Glu	Phe	Phe	Thr		Leu	Asp	Gly	
	CMC	000	210	000		<b></b>		215					220				
											AAA						719
00	447	225	Leu	ALG	ALG	TYL	230	Pro	AIA	Сув	Lys	235	Leu	Leu	Arg	Asp	
20	GAG		ACA	TTC	CAG	GTC		CTC	AAC	CAA	TAT		COO	ccc	ma.	ara.	7.67
											Tyr						767
	240					245	0_1	204	-111	0.111	250	1111	<b>VQI</b>	GIY	ser	255	
25	CTC	CCA	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACG	GTG	GTC	ACC	TCC	АТС		815
											Val						013
					260			_		265					270		
	ACC																818
30	Thr																
	SEQ II																
	SEQUEN							s									
35	SEQUEN					aci	.d										
	STRANI				ole												
	TOPOLO																
	ANTI-S ORIGIN																
40	ORGANI														-		
	IMMEDI																
	CLONE:					300	MCE										
			_														
45	GG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	CTG	47

		His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	Leu	
		1				5					10					15	
5	ATA	GCG	TTC	GCT	TCG	CGG	GGT	AAC	CAC	GTC	TCC	CCC	ACG	CAT	TAT	GTG	95
· ·															Tyr		
					20					25					30		
	CCT	GAG	AGC	GAC	GCC	GCA	GCG	CGT	GTC	ACC	CAG	ATC	CTC	TCC	AGC	СТТ	143
10															Ser		
10				35					40					45			
	ACC	ATC	ACT	CAG	CTG	CTG	AAG	AGG	CTC	CAC	CAG	TGG	ATT	GAT	GAG	GAC	191
															Glu		
			50				-	55					60		014	1101	
15	TGC	TCC	ACG	CCA	TGT	TCT	GGT	TCG	TGG	CTC	AGG	GAT	GTT	TGG	GAC	TGG	239
															Asp		-03
		65					70		_		_	75				E	
	ATA	TGC	ACG	GTG	TTG	AGT	GAC	TTC	AAG	ACC	TGG	CTC	CAG	TCC	AAG	CTC	287
20	Ile	Cys	Thr	Val	Leu	Ser	Asp	Phe	Lys	Thr	Trp	Leu	Gln	Ser	Lys	Leu	
	80					85					90				_	95	
	CTG	CCG	CGG	CTA	CCG	GGA	GTC	ССТ	TTC	CTC	TCA	TGC	CAA	CGT	GGG	TAC	335
															Gly		
25					100					105					110	-	
	AAG	GGA	GTC	TGG	CGG	GGA	GAT	GGC	ATC	ATG	CAG	ACC	ACC	TGC	CCA	TGC	383
															Pro		
				115					120					125		-	
30	GGA	GCA	CAA	ATC	GCC	GGA	CAT	GTC	AAA	AAT	GGT	TCT	ATG	AGG	ATC	ACT	431
															Ile		
			130					135			_		140	_			
	GGC	CCC	AGA	ACC	TGT	AGC	AAC	ACG	TGG	CAC	GGA	ACG	TTC	CCC	ATC	AAC	479
35															Ile		
		145					150	٠				155	•				
	GCG	TAC	ACC	ACA	GGC	CCC	TGC	ACA	CCC	TCC	CCA	GCG	CCG	AAC	TAC	TCC	527
															Tyr		
40	160					165					170					175	
	AGG	GCG	TTA	TGG	CGG	GTA	GCT	GCT	GAG	GAG	TAT	GTG	GAG	GTC	ACG	CGG	575
															Thr		
					180					185	_				190	•	
<b>4</b> 5	GTG	GGG	GAC	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	TTG	AAA	623
															Leu		
						_			-				- &			, -	

				195					200					205			
	TGC	CCA	TGC	CAG	GTC	CCG	GCC	CCC	GAA	TTC	TTC	ACG	GAG	TTG	GAT	GGG	671
																Gly	
5			210					215					220		_	-	
	GTA	CGG	CTA	CAC	AGG	TAC	GCT	CCG	GCG	TGC	AAA	ССТ	CTC	CTA	CGG	GAT	719
																Asp	
	•	225					230					235				_	
10	GAG	GTC	ACA	TTC	CAG	GTC	GGG	CTC	AAC	CAA	TTC	CCG	GTT	GGG	TCG	CAG	767
	Glu	Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Phe	Pro	Val	Gly	Ser	Gln	
	240					245					250			-		255	
	CTC	CCA	TGC	GAG	CCC	GAA	CCG	GAT	GTA	ATA	GTG	GTC	ACC	TCC	ATG	CTC	815
15	Leu	Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Met	Val	Val	Thr	Ser	Met	Leu	
					260					265					270		
	ACC														•		818
	Thr																
20																	
	SEQ I											•					
	SEQUE							cs									
	SEQUE					aci	id						•				
25	STRAN				ole												
	TOPOL													٠			
	ANTI-																
	ORIGI																
30	ORGAN:																
	IMMED:			ERIME	INTAL	SOU	JRCE								-		
	CLONE	H22	2-8														
	66	0 h m	ama	999													
35	GG	CAT															47
		_	vaı	GIA	Pro		Glu	Gly	Ala	Val		Trp	Met	Asn	Arg	Leu	
	ama.	1	mma	000	<b></b>	5					10					15	
		GCG															95
40	ııe	Ala	Pne	Ата		Arg	Gly	Asn	His		Ser	Pro	Thr	His		Val	
70	CCM	CNC	200	a.a	20					25					30		
		GAG															143
	PLO	Glu	ser		ALA	Ala	Ala	Arg		Thr	Gln	Ile	Leu		Ser	Leu	
4E	» CC	አመሮ	3 Cm	35	ama	ama			40					45			
45	ACC	ATC	ACT	CAG	CTG	CTG	AAG	AGG	CTC	CAC	CAG	TGG	ATT	AAT	GAG	GAC	191

			50		Leu			55					60			_	
	TGC	TCC	ACG	CCA	TGT	TCT	GGT	TCG	TGG	CTC	AGG	GAT	GTT	TGG	GAC	TGG	239
5	Cys	Ser	Thr	Pro	Сув	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	Asp	Trp	
		65					70			•		75					
					TTG												287
			Thr	Val	Leu	Ser	Asp	Phe	Lys	Thr	Trp	Leu	Gln	Ser	Lys	Leu	
10	80					85					90					95	
					CCG												335
	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Leu	Ser	Сув	Gln	Arg	Gly	Tyr	
					100					105					110		
15					CGG												383
	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	Gln	Thr	Thr	Cys	Pro	Cys	
				115					120					125			
					GCC												431
20	Gly	Ala		Ile	Ala	Gly	His	Val	Lys	Asn	${\tt Gly}$	Ser	Met	Arg	Ile	Thr	
			130					135					140				
					TGT												479
	Gly		Arg	Thr	Сув	Ser	Asn	Thr	Trp	His	Gly	Thr	Phe	Pro	Ile	Asn	
25		145					150					155					
					GGC												527
		Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	
	160					165					170					175	
30					CGG												575
	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ala	Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	
					180					185					190		
					CAC												623
35	Val	GLY	Asp		His	Tyr	Val	Thr	Gly	Met	Thr	Thr	Asp	Asn	Leu	Lys	
00	mo.a			195					200				•	205			
					GTC												671
•	Cys	Pro		Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	Gly	
40			210					215					220				
40					AGA												719
	val		Leu	His	Arg	Tyr		Pro	Ala	Сув	Lys	Pro	Leu	Leu	Arg	Asp	
	<b></b>	225					230					235					
-					CAG												767
45	G1u	Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Phe	Pro	Val	Gly	Ser	Gln	

											- 7	1 1 1 1 1 1 1	2 ·	1.6			
	240	)				245	<b>i</b> .				250	1000		/		255	
	CTC	CCA	1 TGC	GAG	ccc	GAA	CCG	GAT	GTA	ACA	GTC	GTC	ACC	TCC	ATG	CTC	815
	Lev	Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Thr	Va]	Val	Thr	Ser	Met	Leu	
5					260					265		13.			270		
	ACC				٠						-						818
	Thr	•															
10	SEQ I																
	SEQUE						_	rs									
	SEQUE					c ac	id										
	STRAN				ble												
15	TOPOL																
	ANTI-																
	ORIGI												-				
	ORGAN IMMED																
20	CLONE			EKIM	ENTA	L SO	ORCE										
	CLONE	• 112	2-3														
	GG	САТ	GTG	GGC	CCA	ccc	CAC	CCC	CCM	cmc	020	maa					
			Val														47
25		1		0-1	120	5	GIU	GLY	AIG	vai	10	ттр	Met	ASN	Arg		
	ATA	GCG	TTC	GCT	TCG		GGT	AAC	CAC	CTC		ccc	NCC.	CAM	mam	15	0.5
			Phe														95
					20	-	-			25			****	1110	30	AGI	
30	CCT	GAG	AGC	GAC	GCC	GCA	GCG	CGT	GTC		CAG	ATC	CTC	ሞሮሮ		Cdrds	143
			Ser														143
				35				_	40	•				45			
	ACC	ATC	ACT	CAG	CTG	TTG	AAG	AGG	CTC	CAC	CAG	TGG	ATT		GAT	GAC	191
35			Thr														
			50					55				_	60		-	•	
	TGC	TCC	ACG	CCA	TGT	TCT	GGT	TCG	TGG	CTC	AGG	GAT	GTT	TGG	GAC	TGG	239
			Thr														
40		65					70					75		_	-	•	
	ATA	TGC	ACG	GTG	TTG	AGT	GAC	TTC	AAG	ACC	TGG	CTC	CAG	TCC	AAG	CTC	287
			Thr														
	80					85					90					95	
45	CTG	CCG	CGG	CTA	CCG	GGA	GTC	CCT	TTC	CTC	TCA	TGC	CAA	CGT	GGG	TAC	335

	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Leu	Ser	Cys	Gln	Arg	Gly	Tyr	
					100					105					110		
												ACC					383
5	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	His	Thr	Thr	Cys	Pro	Cys	
				115					120					125			
												TCC					431
	Gly	Ala	Gln	Ile	Ala	Gly	His	Val	Lys	Asn	Gly	Ser	Met	Arg	Ile	Thr	
10			130					135					140				
	GGC	CCC	AGA	ACC	TGT	AGC	AAC	ACG	TGG	CGC	GGA	ACG	TTC	CCC	ATC	AAC	479
	Gly	Pro	Arg	Thr	Cys	Ser	Asn	Thr	Trp	Arg	Gly	Thr	Phe	Pro	Ile	Asn	
		145					150					155					
15	GCG	TAC	ACC	ACA	GGC	CCC	TGC	ACA	CCC	TCC	CCA	GCG	CCG	AAC	TAT	TCT	527
	Ala	Tyr	Thr	Thr	Gly	Pro	Сув	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	
	160					165			٠		170			•		175	
	AAG	GCG	TTG	TGG	CGG	GTA	GCT	GCT	GAG	GAG	TAT	GTG	GAG	GTC	ACG	CGG	575
20	Lys	Ala	Leu	Trp	Arg	Val	Ala	Ala	Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	
					180					185					190	_	
	GTG	GGG	GAT	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	TTG	AAA	623
												Thr					
25				195					200					205		_	
25	TGC	CCA	TGC	CAG	GTC	CCG	GCC	CCC	GAA	TTT	TTC	ACG	GAG	TTG	GAT	GGG	671
	Cys	Pro	Cys	Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	Gly	
			210					215					220				
20	GTA	CGG	CTA	CAC	AGG	TAC	GCT	CCG	GCG	TGC	AAA	ССТ	CTC	CTA	CGG	GAT	719
30	Val	Arg	Leu	His	Arg	Tyr	Ala	Pro	Ala	Cys	Lys	Pro	Leu	Leu	Arg	Asp	
		225					230					235				_	
	GAG	GTC	ACA	TTC	CAG	GTC	GGG	CTC	AAC	CAA	TTC	CCG	GTT	GGG	TCG	CAG	767
	Glu	Val	Thr	Phe	Gln	Val	Gly	Leu	Asn	Gln	Phe	Pro	Val	Gly	Ser	Gln	
35	240					245					250			_		255	
	CTA	CCA	TGC	GAG	ccc	GAA	CCG	GAT	GTA	GCA	GTG	GTC	ACC	TCC	ATG	CTC	815
												Val					
					260				-	265					270		
40	ACC																818
	Thr									•							

SEQ ID NO:81

45 SEQUENCE LENGTH: 311 base pairs

50

	SEQUEN	CE T	YPE:	nuc	leic	aci	đ										
	STRAND	EDNE	SS:	doub	le												
	TOPOLO	GY:	line	ar													
5	ANTI-S	ense	: No														
	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa	titi	s C	viru	s										
	IMMEDI.	ATE :	EXPE	RIME	NTAL	SOU	RCE										
10	CLONE:	N17	-3														
	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	GTC	ACC	TCC	ATG	CTC	ACC	GAC	48
	Cys	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	Thr	Asp	
15	1				5					10			•		15	-	
	CCC	TCC	CAC	ATT	ACA	GCA	GAG	GCG	GCT	AGG	CGT	AGG	CTG	ACC	AGA	GGG	96
						Ala											
				20					25			_		30	_	_	
20	TCT	CCC	CCT	TCC	TCG	ACC	AGT	TCT	TCA	GCT	AGT	CAG	TTG	TCT	GCG	CTT	144
	Ser	Pro	Pro	Ser	Ser	Thr	Ser	Ser	Ser	Ala	Ser	Gln	Leu	Ser	Ala	Leu	
			35					40					45				
	TCT	TCG	CAG	GCA	ACA	TGC	ACT	ACC	CAT	CAG	GGC	GCC	CCA	GAC	ACT	GAC	192
25	Ser	Ser	Gln	Ala	Thr	Сув	Thr	Thr	His	Gln	Gly	Ala	Pro	qaA	Thr	Asp	
		50					55					60				-	
	CTC	ATC	GAG	GCC	AAC	CTC	CTG	TGG	CGG	CAG	GAG	ATG	GGC	GGA	AAC	ATC	240
	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	
30	65					70					75					80	
30	ACC	CGC	GTG	GAG	TCA	GAG	AAC	AAG	ATA	GTA	ATT	CTA	GAC	TCT	TTT	GAA	288
	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	Ile	Val	Ile	Leu	Asp	Ser	Phe	Glu	
					85					90					95		
	CCG	CTT	CGA	GCG	GAG	GAG	GAT	G A									311
35	Pro	Leu	Arg	Ala	Glu	Glu	Asp							•			
				100													
40	SEQ ID																
40	SEQUENC	CE LE	ENGTI	I: 31	ll ba	se p	paire	3									
	SEQUENC					acio	1										
•	STRANDI	EDNES	SS: d	loubl	Le												
	TOPOLOG	Y: 1	inea	ar													
45	ANTI-SE	ENSE:	No														

ORIGINAL SOURCE

	ORGANI	SM:	Hepa	titi	s C	viru	s										
	IMMEDI.	ATE .	EXPE	RIME	NTAL	sou	RCE										
5	CLONE:	N17	-1														
	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	GTC	ACC	TCC	ATG	CTC	ACC	GAC	48
		Glu															
10	1				5					10					15	•	
	CCC	TCC	CAC	ATC	ACA	GCA	GAG	GCG	GCT	AGG	CGT	AGG	CTG	GCC	AGA	GGG	96
	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Arg	Arg	Arg	Leu	Ala	Arg	Gly	
				20					25					30			
15	TCT	CCT	CCT	TCT	TCG	GCC	AGC	TCT	TCA	GCT	AGC	CAG	TTG	TCT	GCG	CCA	144
	Ser	Pro	Pro	Ser	Ser	Ala	Ser	Ser	Ser	Ala	Ser	Gln	Leu	Ser	Ala	Pro	
			35					40	•				45				
		TTG															192
20	Ser	Leu	Lys	Ala	Thr	Сув	Thr	Thr	His	Gln	Asp	Ser	Pro	Asp	Ala	Asp	
		50					55					60					
		ATC															240
		Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	
25	65					70					75					80	
		CGC															288
	Thr	Arg	Val	Glu		Glu	Asn	Lys	Ile	Val	Ile	Leu	Asp	Ser	Ser	Glu	
					85					90					95		
30		CTT -						G A									311
	Pro	Leu	Arg		Glu	Glu	Asp										
				100													
	SEQ ID	NO. C															
35	SEQUENC			¥• 31	1 h=			_									
	SEQUENC					_		•									
	STRANDE					acic	•										
	TOPOLOG																
40	ANTI-SE			••													
	ORIGINA			2													
	ORGANIS				Cv	irus											
	IMMEDIA																

55

45

50

CLONE: N17-2

					CCG												48
	€¥s	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	Thr	Asp	
	1				5					10					15		
5					ACA												96
	Paro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Arg	Arg	Arg	Leu	Thr	Arg	Gly	
				20					25					30			
					TTG												144
10	\$er	Pro		Ser	Leu	Ala	Ser		Ser	Ala	Ser	Gln	Leu	Ser	Ala	Pro	
			35					40					45				
					ACA												192
	ser		гля	Ala	Thr	Cys		Thr	His	His	Asp		Pro	Asp	Ala	Asp	
15	acting	50	~~~				55					60	•				
					AAC												240
		TIE	GIU	АТА	Asn		Leu	Trp	Arg	Gln		Met	Gly	Gly	Asn		
	65	ccc	cmc	CAC	en a	70					75					80	
20					TTA												288
	11 11	Arg	Val	GIU	Leu 85	GIU	ABN	Lys	TTE		TIE	Leu	Asp	Ser		Glu	
	ıCe	Cultur	CGA	GCG	GAG	CAC	Cam	C 3		90					95		
					Glu			GA		•							311
25	0		9	100	GIU	GIU	лър										
				100													
	SEQ ID	NO:8	34														
	SEQUENC			I: 31	ll ba	ase r	oaire	3									
30	SEQUENC					_							•				
	STRAMDI																
	TOP OF O	Y: ]	linea	ar													
	ANTI-SI	ENSE:	No.														
35	ORIGINA	L SC	OURCE	3													
	ORGANIS	M: H	Iepat	itis	C v	irus	5										
	IMM EDIA	ATE E	EXPER	RIMEN	TAL	SOUR	CE										
	CLO)厄:																
40																	
	TST	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	CTC	ACT	TCC	ATG	CTC	ACC	GAC	48
					Pro												
	1				5					10					15	-	
45	CCC	TCC	CAC	ATT	ACA	GCA	GAG	ACG	GCT	AAG	CGT	AGG	CTG	GCC	AGA	GGG	96

	Pro	Ser	His	Ile 20	Thr	Ala	Glu	Thr			Arg	Arg	Leu			Gly	
	ጥርጥ	ccc	CCT		mmc	ccc	»ca	mam	25		. om	~~~		30			
5																CCC	144
	361	FIG	Pro 35	PLO	ьец	ALG	ser		ser	Ala	Ser	Gin			Ala	Pro	
	TICC.	CMC		cac	303	maa	3.00	40					45				
			AAG														192
10	Ser		Lys	AIA	Thr	cys		Thr	His	His	Asp		Pro	Asp	Ala	Asp	
	CEC	50	C2C	000		ama	55					60					
			GAG														240
		TTE	Glu	ALA	Asn		Leu	Trp	Arg	Gln		Met	Gly	Gly	Asn	Ile	
15	65	00m	ama	~~~		70					75					80	
			GTG														288
	Thr	Arg	Val	GIU		GIu	Asn	Lys	Val	Val	Ile	Leu	Asp	Ser	Phe	Asp	
	222	~~~			85		_			90				-	95		
20			CGA					G A									311
	Pro	Leu	Arg		Glu	Glu	Asp										
				100													
	670 TD		~-														
25	SEQ ID																
	SEQUEN					_		5									
•	SEQUEN					acio	i										
	STRANDI		-		le												
30	TOPOLOG			ar													
	ANTI-SI																
	ORIGINA																
	ORGANIS																
35	IMMEDIA			RIMEN	ITAL	SOUF	RCE						•				
	CLONE:	H17-	-3														
	TGT	GAG	CCC	GAA	CCG	GAT	GTA	ACA	GTG	GTC	ACT	TCC	ATG	CTC	ACC	GAC	48
40	Cys	Glu	Pro	Glu	Pro	Asp	Val	Thr	Val	Val	Thr	Ser	Met	Leu	Thr	qaA	
	1				5					10					15		
	CCC	TCC	CAC	TTA	ACA	GCA	GAG	GCG	GCT	GGG	CGT	AGG	CTG	GCC	AGA	GGG	96
	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Gly	Arg	Arg	Leu	Ala	Arg	Gly	
45				20					25					30		_	
	TCT	CCC	CCT	TCC	TTG	GCC	AGC	TCT	TCA	GCT	AGT	CAG	TTG	TCT	GCG	CCC	144
			Pro														

	*																
			35					40					45				
	TCI	CTG	AAG	GCG	ACA	TGC	ACT	ACC	CAT	CAT	GAC	TCC	CCG	GAC	GCT	GAC	192
	Ser	Leu	Lys	Ala	Thr	Cys	Thr	Thr	His	His	Asp	Ser	Pro	Asp	Ala	Asp	
5		50					55					60					
	CTC	ATC	GAG	GCC	AAC	CTC	CTA	TGG	CGG	CAG	GAG	ATG	GGA	GGG	AAC	ATC	240
*	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	
	65					70					75			_	•	80	
10	ACC	CGC	GTG	GAG	TCA	GAG	AGC	AAG	GTA	GTA	ATT	CTG	GAC	TCT	TTC	GAC	288
	Thr	Arg	Val	Glu	Ser	Glu	Ser	Lys	Val	Val	Ile	Leu	Asp	Ser	Phe	Asp	
					85					90					95	-	
	CCG	CTT	CGA	GCG	GAG	GAG	GAT	G A									311
15	Pro	Leu	Arg	Ala	Glu	${\tt Glu}$	Asp										
				100									•		•		
	SEQ ID	NO:	86											•			
20	SEQUEN	CE LI	engti	i: 74	10 ba	se p	paire	3									
	SEQUEN	CE TY	YPE:	nuc	leic	acio	i										
	STRAND	EDNES	SS: c	loub]	le												
	TOPOLO	GY: ]	linea	ır													
25	ANTI-S	ENSE:	No.														
20	ORIGIN	AL S	OURCE	:													
	ORGANI	SM: I	lepat	itis	3 C v	irus	5										
	IMMEDIA	ATE E	XPER	IMEN	ITAL	SOUF	CE										
	CLONE:	028-	-1														
30																	
	GTG	GTA	GTC	CTG	GAC	TCG	TTG	GAG	CCG	CTT	CAA	GCG	AAG	GAA	GGT	GAG	48
		Val															
	1				5					10			-		15		
35	AGG	GAA	GTG	TCC	GTT	GCG	GCG	GAG	ATC	CTG	CGG	AAG	ACC	AGG		TTC	96
	Arg	Glu	Val	Ser	Val	Ala	Ala	Glu	Ile	Leu	Arg	Lys	Thr	Ara	Lvs	Phe	
				20					25		•	_		30	-1-	0	•
	ccc	GCA	GCG	ATG	CCC	GTA	TGG	GCA	CGC	CCG	GAC	TAC	AAC		CCA	ጥጥል	144
40		Ala															
			35				•	40	,		<i>E</i>		45				
	CTA	GAG	TCT	TGG	AAG	AAC	CCG	-	TAC	GTC	CCT	CCA		GTA	CAC	ഭദദ	192
		Glu															172
45		50		_	_		55	•			- <del>-</del>	60				1	

	TGC	CCA	TTG	CCG	CCT	ACC	AAG	GCC	CCT	CCA	ATA	CCA	CCT	CCA	CGA	AGA	240
		Pro															
	65					70					75				•	80	
5	AAG	AGA	ACG	GTT	GTC	CTG	ACA	GAA	TCC	TCC	GTG	TCC	TCT	GCC	TTG	GCG	288
		Arg															
					85					90					95		
	GAG	CTT	GCT	ACA	AAG	ACC	TTT	GGC	AGT	TCC	GGA	TCG	TCG	GCC	GTC	GAC	336
10		Leu															
				100					105					110		•	
	AGC	GGC	ACG	GCG	ACC	GGC	CCT	ССТ	GAC	CAG	GCC	TCC	GCC	GAA	GGA	GAT	384
		Gly															
15			115					120					125		_	-	
	GCA	GGA	TCC	GAC	GCT	GAG	TCG	TAC	TCC	TCC	ATG	CCC	CCC	CTT	GAG	GGA	432
		Gly															
		130					135					140				_	
20	GAG	CCG	GGG	GAC	CCC	GAT	CTC	AGC	GAC	GGG	TCT	TGG	TCT	ACC	GTA	AGC	480
	Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser	
	145					150		•			155	_				160	
	GAG	GAG	GCC	AGC	GAG	GAC	GTC	GTC	TGC	TGC	TCG	ATG	TCC	TAC	ACA	TGG	528
25		Glu															
					165					170					175	-	
	ACA	GGC	GCC	TTA	ATT	ACA	CCA	TGC	GCC	GCG	GAG	GAG	AGC	AAG	CTG	CCC	576
	Thr	Gly	Ala	Leu	Ile	Thr	Pro	Cys	Ala	Ala	Glu	Glu	Ser	Lys	Leu	Pro	
30				180					185					190			
	ATT	AAT	GCG	CTG	AGC	AAC	CCT	TTG	CTG	CGC	CAC	CAC	AAC	ATG	GTC	TAT	624
	Ile	Asn	Ala	Leu	Ser	Asn	Pro	Leu	Leu	Arg	His	His	Asn	Met	Val	Tyr	
			195					200					205				
35	GCC	ACA	ACA	TCC	CGC	AGC	GCA	AGC	CAG	CGG	CAG	AAA	AAG	GTC	ACA	TTT	672
	Ala	Thr	Thr	Ser	Arg	Ser	Ala	Ser	Gln	Arg	Gln	Lys	Lys	Val	Thr	Phe	
		210					215					220				•	
	GAC	AGA	CTG	CAA	GTC	CTG	GAT	GAC	CAC	TAC	CGG	GAC	GTG	CTC	AAG	GAC	720
40	Asp	Arg	Leu	Gln	Val	Leu	Asp	Asp	His	Tyr	Arg	Asp	Val	Leu	Lys	Asp	
70	225					230					235					240	
	ATG	AAG	GCC	AAG	GCG	TCC	AC										740
	Met	Lys	Ala	Lys	Ala	Ser											
					245												

	SEQ ID NO:87
	SEQUENCE LENGTH: 740 base pairs
	SEQUENCE TYPE: nucleic acid
5	STRANDEDNESS: double
	TOPOLOGY: linear
	ANTI-SENSE: NO
	ORIGINAL SOURCE
10	ORGANISM: Hepatitis C virus
	IMMEDIATE EXPERIMENTAL SOURCE
	CLONE: 028-2

15	GTG	GTA	GTC	CTG	GAC	TCG	TTG	GAC	CCG	CTT	CGA	GCG	GAG	GAA	GAT	GAG	48
	Val	Val	Val	Leu	Asp	Ser	Leu	Asp	Pro	Leu	Arg	Ala	Glu	Glu	Asp	Glu	
	1				5					10					15		
	AGG	GAA	GTG	TCC	GTT	GCG	GCG	GAG	ATC	CTG	CGA	AAG	ACC	AAG	AAA	TTC	96
20						Ala											
				20					25					30			
	CCC	GCA	GCG	ATG	CCC	GTA	TGG	GCA	CGC	CCG	GAC	TAC	AAC	CCA	CCA	TTA	144
	Pro	Ala	Ala	Met	Pro	Val	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	
25			35					40					45				
	CTA	GAG	TCT	TGG	AAG	AAC	CCG	GAC	TAC	GTC	CCT	CCG	GTG	GTA	CAC	GGG	192
	Leu	Glu	Ser	Trp	Lys	Asn	Pro	Asp	Tyr	Val	Pro	Pro	Val	Val	His	Gly	
		50					55					60					
30						ACC											240
•	Cys	Pro	Leu	Pro	Pro	Thr	Lys	Ala	Pro	Pro	Ile	Pro	Pro	Pro	Arg	Arg	
	65					70					75					80	
						CTG											288
35	Lys	Arg	Thr	Val	Ala	Leu	Thr	Glu	Ser	Thr	Val	Ser	Ser	Ala	Leu	Ala	
30					85					90					95		
						ACC											336
-	Glu	Leu	Ala	Thr	Lys	Thr	Phe	Gly	Ser	Ser	Gly	Ser	Ser	Ala	Val	Asp	
				100					105					110			
40						GGC											384
	Ser	Gly	Thr	Ala	Thr	Gly	Pro	Pro	Asp	Gln	Ala	Ser	Ala	Glu	Gly	Asp	
			115					120					125				
						GAG											432
45	Ala	Gly	Ser	Asp	Ala	Glu	Ser	Tyr	Ser	Ser	Met	Pro	Pro	Leu	Glu	Gly	

		130					135					140					
	GAG	CCG	GGG	GAC	CCT	GAT	CTC	AGC	GAC	GGG	TCT	TGG	TCT	ACT	GTA	AGC	480
		Pro															
5	145					150					155	_				160	
	GAG	GAG	GCC	GGC	GAG	GAC	GTC	GTC	TGC	TGC	TCG	ATG	TCC	TAC	ACA	TGG	528
	Glu	Glu	Ala	Gly	Glu	Asp	Val	Val	Cys	Cys	Ser	Met	Ser	Tyr	Thr	Trp	
	•				165					170					175		
10		GGC															576
	Thr	Gly	Ala	Leu	Ile	Thr	Pro	Сув	Ala	Ala	Glu	Glu	Ser	Lys	Leu	Pro	
				180					185					190			
		AAT															624
15	Ile	Asn		Leu	Ser	Asn	Ser	Leu	Leu	Arg	His	His	Asn	Met	Val	Tyr	
			195					200					205				
		ACA															672
	AIa	Thr	Thr	Ser	Arg	Ser		Ser	Gln	Arg	Gln		Lys	Val	Thr	Phe	
20	CAC	210	ama.		-		215					220					
		AGA															720
	225	Arg	nen	GIN	vai		Asp	Asp	His	Tyr		Asp	Val	Leu	Lys		
		AAG	ccc	A A C	ccc	230	20				235					240	
25		Lys					AC										740
			mia	nys	245	Ser											
					213												
	SEQ ID	NO:8	8									-					
30	SEQUENC			I: 74	0 ba	se r	airs										
	SEQUENC					_											
	STRANDE	DNES	S: d	loubl	.e												
	TOPOLOG	Y: 1	inea	r													
35	ANTI-SE	NSE:	No														
	ORIGINA	L SO	URCE														
	ORGANIS	м: н	epat	itis	Cv	irus	i										
	IMMEDIA	TE E	XPER	IMEN	TAL	SOUR	CE										
40	CLONE:	028-	4														
	GTG	GTA	GTC	CTG	GAC	TCG	TTG	GAC	CCG	CTT	CGA	GCG	GAG	GAA	GAT	GAG	48
		Val															
45	1				5					10					15		

	AGG	GAA	GTG	TCC	GTT	GCG	GCG	GAG	ATC	CTG	CGA	AAG	ACC	AAG	AAA	TTC	96
												Lys					
				20					25			_		30	•		
5	CCC	GCA	GCG	ATG	CCC	GTA	TGG	GCA	CGC	CCG	GAC	TAC	AAC	CCA	CCA	TTA	144
	Pro	Ala	Ala	Met	Pro	Val	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	
			35					40					45				
	CTA	GAG	TCT	TGG	AAG	AAC	CCG	GAC	TAC	GTC	CCT	CCG	GTG	GTA	CAC	GGG	192
10	Leu	Glu	Ser	Trp	Lys	Asn	Pro	Asp	Tyr	Val	Pro	Pro	Val	Val	His	${ t Gly}$	
		50					55					60					
												CCA					240
		Pro	Leu	Pro	Pro		Lys	Ala	Pro	Pro	Ile	Pro	Pro	Pro	Arg	Arg	
15	65					70					75					80	
												TCC					288
	гуз	Arg	Thr	Val		Leu	Thr	Glu	Ser		Val	Ser	Ser	Ala	Leu	Ala	
	CAC	com	CCM	202	85	300	mmm			90					95		
20												TCG					336
	GIU	nea	AIG	100	пåр	THE	Phe	GIY	105	ser	GIY	Ser	ser		Val	qaA	
	AGC	GGC	ACG		ACC	GGC	ር ር	_{ССФ}		CAG	ccc	TCC	CCC	110	CCA	C3 M	204
												Ser					384
25		-	115			1		120	p	01	1114	Der	125	GLU	GLY	лар	
	GCA	GGA	TCC	GAC	GCT	GAG	TCG	TAC	TCC	TCC	ATG	ccc	_	CTT	GAG	GGA	432
												Pro					
		130					135					140				-	
30	GAG	CCG	GGG	GAC	CCT	GAT	CTC	AGC	GAC	GGG	TCT	TGG	TCT	ACT	GTA	AGC	480
	Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser	
	145					150					155					160	
05												ATG					528
35	Glu	Glu	Ala	Gly	Glu	Asp	Val	Val	Cys	Cys	Ser	Met	Ser	Tyr	Thr	Trp	
					165					170					175		
												GAG					576
40	Thr	GLY	Ala		Ile	Thr	Pro	Сув		Ala	Glu	Glu	Ser	Lys	Leu	Pro	
70	» mm	3 3 M	aaa	180					185					190			
												CAC					624
	TIG	asn		ren	ser	ASN	ser		Leu	Arg	His	His		Met	Val	Tyr	
45	GCC	אכא	195	mcc	ccc	N.C.C	cas	200	<b>63.6</b>		a		205				
40	GCC	ACA	ACA	TCC	CGC	AGC	GCA	AGC	CAG	CGG	CAG	AAA	AAG	GTC	ACA	TTT	672

	Al	a Th	r Thi	Sei	Arc	j Ser	A 1.15 h		r Gl	n Ar	g Glr	Lys	Lye	Va	l Thr	Phe	
	GA			2 CA2	Carc	· Cmc	215					220			- ''	5	
5	As	p Ar	A CTO	ı Glr	Val	Len	, gai	Acr	CA	C TA	C CGG	GA	GTO	CTC	AAG	GAC	720
• .	. 22	F	g Lei	- 011		230		, wel	, HT	в ту:			Val	Let	ı Lys		•
			G GCC	. AAG	GCG						235	•		17.4		240	
			s Ala									1				٧.	740
10		_		1-	245		•										
	SEQ I	D NO	:89														
	SEQUE	NCE	LENGI	H: 5	15 b	ase	pair	s									
15	SEQUE	NCE	TYPE:	nuc	leic	aci	ď										
	STRAN	DEDN	ESS:	doub	le												
	TOPOL	OGY:	line	ar									•				
	ANTI-	SENS	E: No	)													
20	ORIGI	NAL	SOURC	E													
	ORGAN																
	IMMED			RIME	NTAL	SOU	RCE										
	CLONE	: N2	9-1														
25																	
	AC		CGG														47
			Arg	Asp	Val :		Lys	Glu	Met	Lys	Ala	Lys	Ala	Ser	Thr V	Val	
		1				5					10					15	
30	AAC	GC!	T AAA -	CTT	CTA	TCT	GTA	GAG	GAA	GCC	TGC	AAG	CTG	ACG	CCC	CCA	95
	гА	3 Ali	a Lys	Leu		Ser	Val	Glu	Glu	Ala	Cys	Lys	Leu	Thr	Pro	Pro	
	CN	. ma		202	20					25					30		
	CAC	. 200	G GCC	AGA	TCT	AAA	TTT	GGC	TAC	GGG	GCA	AAG	GAC	GTC	CGG	AGC	143
35	nıs	) Se.	r Ala		ser	гля	Phe	GLY			Ala	Lys	Asp	Val	Arg	Ser	
	ርሞር	2 ጥርረ	י אכני	35	CCC	C C C C C C C C C C C C C C C C C C C	330		40					45			
	Tes	9 100	AGC	Tara	315	GII	AAC	CAC	ATC	CGC	TCC	GTG	TGG	AAG	GAC	TTG	191
			ser 50	ыyь	VIG	Vai	ASII		TTE	Arg	Ser	Val		Lys	Asp	Leu	
40	СТС	GAZ		ልሮሞ	CAC	ארא	CCX	55 200	CNO	300			60				
	Lei	Gla	A GAC	Thr	GAG	mh~	Dro	ATT	GAC	ACC	ACC	ATC	ATG	GCA	AAA	AAT	239
	200	65	ı Asp	1111	GIU	Ant	70	116	Авр	Thr	Thr		Met	Ala	Lys	Asn	
	GAG			ጥረጥ	ርጥጥ	CAA		CNC		221		75					
45	Glu	. Val	TTC	CAR	Val	Gln	Dro	Clu	AAA	GGA	GGC	CGC	AAG	CCA	GCT	CGC	287
-	<b></b>		l Phe	~ <u>1</u> 5	141	GIII	FIO	GIU	гЛя	стА	стА	Arg	тАв	PLO	Ala	Arg	
50																	

	80					85					90					95	
	CTT	ATC	GTA	TTC	CCA		TTG	GGG	CTTT	ሮሮጥ		ጥርር	GAG	222	y days.	GCC	335
		Ile															333
5					100	•		4		105		012	-	270	110	mu	
	CTC	TAC	GAC	GTG	GTC	TCC	ACT	СТТ	CCT	CAG	GCC	GTG	ATG	GGC		TCA	383
		Tyr															555
	•			115					120					125			
10	TAC	GGA	TTC	CAG	TAC	TCC	CCT	GGA	CAG	CGG	GTC	GAG	TTC	CTG	GTG	AAT	431
		Gly															
			130					135					140				
	GCC	TGG	AAG	TCA	AAG	AAG	AGC	CCT	ATG	GGC	TTT	GCA	TAT	GAC	ACC	CGC	479
15	Ala	Trp	Lys	Ser	Lys	Lys	Ser	Pro	Met	Gly	Phe	Ala	Tyr	Asp	Thr	Arg	
		145					150					155					
		TTT															515
		Phe	qaA	Ser	Thr	Val	Thr	Glu	Asn	Asp	Ile	Arg					
20	160					165					170						
	SEQ ID																
	SEQUEN							5									
25	SEQUEN					acio	i										
	STRANDI				Le												
	TOPOLOG			ır													
	ORIGINA	-		7													
30	ORGANIS					. ł											
	IMMEDIA																
	CLONE:			LITTE	ITAL	SOUR	CE										
	0_02.	1123	-														
<b>3</b> 5	AC 1	TAC C	ec c	AC G	ምር ሰ	'ጥር' አ	AC 0	יאכ א	መሮ አ	3C C	'CC 3					ımm	
		lyr A															47
		1	•			5	.,				10	iyo v	та 5	et i	.ILL V	15	
	AAG	GCT	AAA	CTT	CTA	_	GTA	GAG	GAA			AAG	ርሞር	ACG	CCC		95
40		Ala															33
					20					25	-1-	-1-	٠,٠		30	110	
	CAC	TCG	GCC	AGA	TCT	AAA	TTT	GGC	TAC	GGG	GCA	AAG	GAC	GTC		AGC.	143
		Ser															110
<b>4</b> 5				35				_	40	-		•	•	45			
														-			
50																	

						GTT											191
	Leu	Ser	Ser 50	Lys	Ala	Val	Asn	His 55	Ile	Arg	Ser	Val		Lys	Asp	Leu	
5	CTG	CAA		ልርጥ	GNG	ACA	CCA		CAC	200	».cc	» ma	60	CON		220	220
						Thr											239
	200	65			014	****	70	110	nap	THE	TIIL	75	Mec	MIG	пув	Abii	
	GAG		TTC	TGT	GTT	CAA		GAG	AAA	GGA	ccr		AAC	CCA	CCT	cec	287
10						Gln											201
	80			-		85				2	90	5	-,,-			95	
	CTT	ATC	GTA	TTC	CCA	GAC	TTG	GGG	GTT	CGT	GTG	TGC	GAG	AAA	ATG		335
						Asp											
15					100					105				_	110		
	CTC	TAC	GAC	GTG	GTC	TCC	ACT	CTT	CCT	CAG	GCC	GTG	ATG	GGC	TCC	TCA	383
	Leu	Tyr	Asp	Val	Val	Ser	Thr	Leu	Pro	Gln	Ala	Val	Met	Gly	Ser	Ser	
				115					120					125			
20						TCC											431
	Tyr	Gly		Gln	Tyr	Ser	Pro	Gly	Gln	Arg	Val	Glu	Phe	Leu	Val	Asn	
			130					135					140				
						AAG											479
25	Ala		Lys	Ser	Lys	Lys		Pro	Met	Gly	Phe	Ala	Tyr	qaA	Thr	Arg	
		145					150					155		•			
						GTC											515
		Pne	Asp	Ser	Thr	Val	Thr	Glu	Asn	Asp		Arg					
30	160					165					170						
	SEQ ID	NO: 9	31														
	SEQUENC			I: 50	)3 ba	se r	a i rs										
	SEQUENC					_											
35	STRANDE	DNES	SS: d	loubl	Le							-					
	TOPOLOG	Y: ]	linea	ır													
	ANTI-SE	NSE:	No:														
	ORIGINA	L SC	OURCE	3													
40	ORGANIS	M: F	lepat	itis	CV	rirus	;										
	IMMEDIA	TE I	EXPER	RIMEN	ITAL	SOUF	CE						•				
	CLONE:	N29-	-3														

AC TAC CGG GAC GTG CTG AAG GAG ATG AAG GCG AAG GCG TCC ACA GTT 47

		Tyr	Arg	Asp	Val	Leu	Lys	Glu	Met	Lys .	Ala :	Lys .	Ala :	Ser '	Thr	Val	
		1				5					10					15	
	AAG	GCT	AAA	CTT	CTA	TCT	GTA	GAG	GAA	GCC	TGT	AAG	CTG	ACG	CCC	CCA	95
5	Lys	Ala	Lys	Leu	Leu	Ser	Val	Glu	Glu	Ala	Сув	Lys	Leu	Thr	Pro	Pro	
					20					25					30		
	CAC	TCG	GCC	AGA	TCT	AAG	TTT	GGC	TAC	GGG	GCA	AAG	GAC	GTC	CGG	AGC	143
	His	Ser	Ala	Arg	Ser	Lys	Phe	Gly	Tyr	Gly	Ala	Lys	Asp	Val	Arg	Ser	
10				35					40					45			
	CTG	TCC	AGC	AAG	GCC	GTT	AAC	CAC	ATC	CGC	TCC	GTG	TGG	AGG	GAC	TTG	191
	Leu	Ser	Ser	Lys	Ala	Val	Asn	His	Ile	Arg	Ser	Val	Trp	Glu	Asp	Leu	
			50					55					60				
15				ACT													239
	Leu	Glu	Asp	Thr	Glu	Thr	Pro	Ile	Asp	Thr	Thr	Ile	Met	Ala	Lys	Asn	
		65					70					75 [.]					
				TGT													287
20		Val	Phe	Cys	Val		Pro	Glu	Lys	Gly	Gly	Arg	Lys	Pro	Ala	Arg	
	80					85					90					95	
				TTC													335
	Leu	Ile	Val	Phe		Asp	Leu	Gly	Val		Val	Сув	Glu	Lys	Met	Ala	
25	~~~				100					105					110		
				GTG													383
	rea	ıyr	Asp	Val	Val	Ser	Thr	Leu		Gln	Ala	Val	Met		Ser	Ser	
	ma C	CCA	mma	115	m. 0	maa	~~m		120					125			
30				CAG													431
	TAT	GIY	130	Gln	TYL	ser	Pro		GIN	Arg	Val	GIu		Leu	Val	Asn	
	GCC	TCC		መሮ አ	AAC	220	a.cm	135	» mc	000	anna.	<b></b>	140	a. a			450
				TCA													479
35		145	ny o	Ser	цуз	цуъ	150	PIO	met	GIY	Pne		туг	Asp	Thr	Arg	
	ጥርጥ		GAC	TCA	»CG	ርሞር		CAC				155					503
				Ser													503
	160	- 110	ımp	OCI	1111	165	1111	GIU									
40	200					100											
	SEQ ID	NO:	92														
	SEQUENC			H: 41	)1 ha	ise r	naire										
	SEQUENC							•									
45							-										

STRANDEDNESS: double

TOPOLOGY: linear

	ANTI-S	SENSE	: No	)													
	ORIGI	NAL S	OURC	E													
5	ORGAN	ISM:	Hepa	titi	s C	viru	s										
	IMMED	IATE	EXPE	RIME	NTAL	SOU	RCE										
	CLONE	: N18	3-4														
10	TG	GGG	ATC	CCG	TAT	GAT	ACC	CGC	TGC	TTT	GAC	TCA	ACG	GTC	ACT (	GAG	47
		Gly	Ile	Pro	Tyr	Asp	Thr	Arg	Cys	Phe	Asp	Ser	Thr	Val	Thr (	Glu	
		1				5					10					15	
	AAT	r GAC	ATC	CGT	ACT	GAG	GAG	TCA	TTA	TAT	CAA	TGT	TGT	GAC	TTG	GAC	95
15	Asr	ı Asp	Ile	Arg	Thr	Glu	Glu	Ser	lle	Tyr	Gln	Cys	Сув	Asp	Leu	Asp	
					20					25					30		
															TAT		143
•	Pro	Glu	Ala	Arg	Gln	Ala	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	
20				35					40				-	45			
															CGC		191
	Gly	7 Gly			Thr	Asn	Ser	Lys	Gly	Gln	Asn	Сув	Gly	Tyr	Arg	Arg	
			50					55					60				
25															CTC		239
	Суа			Ser	Gly	Val			Thr	Ser	Cys	Gly	Asn	Thr	Leu	Thr	
	ma=	65					70					75					
															CAG		287
30			Leu	гÀг	Ala		Ala	Ala	Сув	Arg			Lys	Leu	Gln	Asp	
	80		<b>.</b>	ama	<b>a</b> ma	85					90					95	
															GAA		335
	Cys	THE	met	ren		Сув	GIY	Asp	Asp		Val	Val	Ile	Суѕ	Glu	Ser	
35	ccc	CCN	NCC.	CAC	100	C2.C	000	003		105					110		
															GAG		383
	71.10	. Gry	1111	115	GIU	Asp	AIG	WIG		теп	Arg	val	Pne		Glu	Ala	
	ΔͲር	አቦሮ	AGG		መርር	ccc			120					125			401
40			Arg														401
	1100		130	Aon	DCI	Ald							•				
			130										•				
	SEQ ID	NO:	93														
45	SEQUEN	ICE L	ENGT	H: 4	01 ba	ase	pair	8									

50

	SEQUEN	ICE :	TYPE:	nuc	cleid	aci	.d										
	STRAND	EDNI	ess:	doub	ole												
	TOPOLO	GY:	line	ear													
5	ANTI-S	ENSI	2: No	)													
	ORIGIN	IAL S	SOURC	CE.													
	ORGANI	SM:	Нера	atiti	s C	viru	s										
	IMMEDI	ATE	EXPE	ERIME	NTAL	sou	RCE										
10	CLONE:	N18	3-2														
	TG	GGG	ATC	CCG	TAT	GAT	ACC	CGC	TGC	TTT	GAC	TCA	ACA	GTC	ACT (	GAG	47
		Gly	Ile	Pro	Tyr	Asp	Thr	Arg	Cys	Phe	Asp	Ser	Thr	Val	Thr (	Glu	
15		1				5					10					15	
															TTG		95
	Asn	yst	Ile	Arg	Ile	Glu	Glu	Ser	Ile	Tyr	Glr	Сув	Сув	Asp	Leu	Val	
					20					25					30		
20															TAT		143
	Pro	Glu	Ala			Ala	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	
				35					40					45			
															CGC		191
25	Gly	Gly			Thr	Asn	Ser	Lys	Gly	Gln	Asn	Сув	Gly	Tyr	Arg	Arg	
			50					55		•			60				
															CTC		239
	Cys			Ser	Gly	Val		Thr	Thr	Ser	Сув		Asn	Thr	Leu	Thr	
30	mam	65					70					75					
															CGG		287
	698 80	туг	ьеu	гÃ2	ATa		Ala	Ala	Cys	Arg			Lys	Leu	Arg	_	
		N.C.C	እመረግ	CMC	ama	85 mag	001				90					95	
35															GAA		335
	Cys	1111	Met	ren	100	Cys	GIA	Asp	Asp		Val	Val	He	Cys	Glu	Ser	
	GCG	ccc	NCC	CAC		CNC	<b></b>	221		105				:_	110		
															GAG		383
40	*****	GLY	1111	115	GIU	Asp	AId	WTG		Leu	Arg	vaı	Pne		Glu	Ala	
	ATC	ACC	AGG		ሞርር	ecc			120				•	125			
			Arg														401
			130		OGI	TTO											
45			-50														

SEQ ID NO:94

	SECOEM	CE L	ENGT	n: 4	OT D	ase	pair	s									
	SEQUEN	CE T	YPE:	nuc	leic	aci	d										
5	STRAND	EDNE	ss:	doub	le								-				
	TOPOLO	GY:	line	ar													
	ANTI-S	ENSE	: No														
	ORIGIN	AL S	OURC	E													
10	ORGANI	SM:	Нера	titi:	s C	viru	s									•	
	IMMEDI	ATE	EXPE	RIME	NTAL	SOU	RCE										
	CLONE:	N18	-3														
15	TG	GGG .	ATC (	CCG !	TAT	GAT A	ACC (	CGC	TGC	TTT	GAC	TCA	ACG	GTC	ACT	GAG	47
		Gly	Ile :	Pro !	Tyr .	Asp :	Thr .	Arg	Cys	Phe	Asp	Ser	Thr	Val	Thr	Glu	
		1				5					10					15	
	AAT	GAC	ATC	CGT	ACT	GAG	GAG	TCA	ATT	TAT	CAA	TGT	TGT	GAC	TTG	GAC	95
20	Asn	Asp	Ile	Arg	Thr	Glu	Glu	Ser	Ile	Tyr	Gln	Сув	Cys	Asp	Leu	Asp	
					20					25					30	_	
	CCC	GAG	GCC	AGA	CAG	GCC	ATA	AGG	TCG	CTC	ACA	GAG	CGG	CTT	TAT	ATC	143
	Pro	Glu	Ala	Arg	Gln	Ala	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	
25				35					40					45			
	GGG	GGC	CCC	TTG	ACC	AAT	TCA	AAA	GGG	CAG	AAC	TGC	GGT	TAT	CGC	CGG	191
	Gly	Gly	Pro	Leu	Thr	Asn	Ser	Lys	Gly	Gln	Asn	Сув	Gly	Tyr	Arg	Arg	
			50					55					60				
30	TGC	CGC	GCC	AGC	GGC	GTG	CTG	ACG	ACT	AGC	TGC	GGT	AAT	ACC	CTT	ACA	239
	Cys	Arg	Ala	Ser	Gly	Val	Leu	Thr	Thr	Ser	Сув	Gly	Asn	Thr	Leu	Thr	
		65					70			÷		75					
	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	TGT	CGA	GCT	GCG	AAG	CTC	CAG	GAC	287
35	Суз	Tyr	Leu	Lys	Ala	Ser	Ala	Ala	Cys	Arg	Ala	Ala	. Lys	Leu	Gln	Asp	
••	80					85					90					95	
						TGC											335
	Суз	Thr	Met	Leu	Val	Cys	Gly	Asp	Asp	Leu	Val	Val	Ile	Cys	Glu	Ser	
40					100					105					110		
						GAC											383
	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	Asn	Leu	Arg	Val	Phe	Thr	Glu	Ala	
				115					120					125			
45				TAA										-			401
45	Met	Thr	Arg	Asn	Ser	Ala											

55

	SEQ ID	NO:	95														
5	SEQUEN	ICE L	ENGT	H: 4	01 b	ase	pair	cs									
	SEQUEN	CE T	YPE:	nuc	leic	aci	.d										
	STRAND	EDNE	ss:	doub	le												
	TOPOLO	GY:	line	ar													
10	ANTI-S	ENSE	: No														
	ORIGIN	AL S	OURC	E													
	ORGANI	SM:	Hepa	titi	вC	viru	ıs										
	IMMEDI																
15	CLONE:									-							
15			•														
	TG	GGG .	ATC ·	CCG	TAT	GAT	ACC	CGC	TGC	TTT	GAC	TCA	ACA	GTC	ACT	GAG	47
															Thr (		4,
		1				5		•	-		10					15	
20	AGT	GAT	ATC	CGT	GTT	GAG	GAG	TCA	ATC	TAC	CAA	TGT	TGT	GAC	TTG		95
															Leu		
					20					- 25		•			30		
	ccc	GAG	GCC	AGA	CAG	GCC	АТА	AGG	TCG	CTC	ACA	GAG	CGG	CTT	TAT	ATC	143
25															Tyr		
				35					40				•	45			
	GGG	GGC	CCC	CTG	ACT	AAT	TCA	AAA	GGG	CAG	AAC	TGC	GGT	TAT	CGC	CGG	191
															Arg		
30			50					55					60	_	_	_	
	TGC	CGC	GTC	AGC	GGC	GTG	CTG	ACG	ACC	AGC	TGC	GGT	AAT	ACT	CTT	ACA	239
	Сув	Arg	Val	Ser	Gly	Val	Leu	Thr	Thr	Ser	Cys	Gly	Asn	Thr	Leu	Thr	
		65				•	70					75					
35	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	TGT	CGA	GCT	GCA	AAG	CTC	CAG	GAC	287
	Cys	Tyr	Leu	Lys	Ala	Ser	Ala	Ala	Сув	Arg	Ala	Ala	Lys	Leu	Gln	Asp	
	80					85					90					95	
	TGC	ACA	ATG	CTC	GTG	TGC	GGG	GAC	GAC	CTT	GTC	GTC	ATC	TGT	GAG	AGC	335
40	Суз	Thr	Met	Leu	Val	Cys	Gly	Asp	Asp	Leu	Val	Val	Ile	Сув	Glu	Ser	
					100					105			•		110		
	GCG	GGA	ACC	CAG	GAG	GAC	GCG	GCG	AAC	CTA	CGA	GTC	TTC	ACG	GAG	GCT	383
	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	Asn	Leu	Arg	Val	Phe	Thr	Glu	Ala	
45				115					120					125			

1 1	ATG ACC AGG	AAT TCC GCC		1.	*			401			
	Met Thr Arg	Asn Ser Ala									
	130										
5											
	SEQ ID NO:96					•					
	SEQUENCE LENGTH	: 401 base	pairs								
	SEQUENCE TYPE:		_								
10	STRANDEDNESS: double										
	TOPOLOGY: linear										
	ANTI-SENSE: NO										
	ORIGINAL SOURCE	:									
15	ORGANISM: Hepat	itis C viru	s								
13	IMMEDIATE EXPER										
	CLONE: H18-2										
	•										
00	TG GGG ATC C	CG TAT GAT	ACC CGC	TGC TTT	GAC TCA A	ACA GTC	ACT GAG	47			
20		ro Tyr Asp !									
	1	5			10		15				
	AGT GAT ATC	CGT GTT GAG	GAG TCA	ATC TAC	CAA TGT	TGT GAC	TTG GCC	95			
05	Ser Asp Ile	Arg Val Glu	Glu Ser	Ile Tyr	Gln Cys	Cys Asp	Leu Ala				
25		20		25			30				
	CCC GAG GCC	AGA CAG GCC	ATA AGG	TCG CTC	ACA GAG	CGG CTT	TAT ATC	143			
	Pro Glu Ala	Arg Gln Ala	Ile Arg	Ser Leu	Thr Glu	Arg Leu	Tyr Ile				
		35		40		45					
30	GGG GGC CCC	CTG ACT AAT	TCA AAG	GGG CAG	AAC TGC	GGT TAT	CGC CGG	191			
	Gly Gly Pro	Leu Thr Asn	Ser Lys	Gly Gln	Asn Cys	Gly Tyr	Arg Arg				
	50		55			60					
	TGC CGC GTC							239			
35	Cys Arg Val	Ser Gly Val	Leu Thr	Thr Ser	Cys Gly	Asn Thr	Leu Thr				
	65		70		75						
	TGT TAC TTG							287			
	Cys Tyr Leu	Lys Ala Ser	Ala Ala	Cys Arg	Ala Ala	Lys Leu	Gln Asp				
40	80	85			90		95				
	TGC ACA ATG							335			
	Cys Thr Met		Gly Asp		Val Val	Ile Cys	Glu Ser				
		100		105			110				
45	GCG GGA ACC	CAG GAG GAC	GCG GCG	AAC CTA	CGA GTC	TTC ACG	GAG GCT	383			
50											

5	YTA	G AC	y Thr C AGG r Arg 130	115 AAT Asn	r TCC	GCC		Ala	120		ı Arç	y Val	. Phe	Th:		ı Ala	401
10	SEQ II SEQUEI SEQUEI STRANI	NCE	LENGT TYPE:	nuc	leic		_	8									
15	TOPOLO ANTI-S ORIGIN ORGANI IMMEDI	SENS NAL : ISM:	E: No SOURC Hepa	E titi					-					•			
20	CLONE	H1	8-3														
	TG	Gly	ATC Ile			Asp '					Asp						47
25			T ATC														95
30			G GCC														143
			C CCC							CAG					CGC		191
35			50 C GTC					55					60				000
			y Val														239
40			TTG														287
	80 80		. Leu	тÃ2	Ala	Ser 85	Ala	ATA	Cys	Arg	Ala 90	Ala	Lys	Leu	Gln	Asp 95	
45			ATG														335
70	Cya	1111	Met	nea	val	Cys	GTA	Авр	Asp	ren	val	val	TTE	Cys	Glu	ser	

					100					105					110		
	GCG	GGA	ACC	CAG	GAG	GAC	GCG	GCG	AAC	СТА	CGA	GTC	TTC	ACG	GAG	GCT	383
																Ala	
5				115					120		•			125			
	ATG	ACC	AGG	AAT	TCC	GCC											401
	Met	Thr	Arg	Asn	Ser	Ala											101
	•		130														
10																	
	SEQ ID	NO:	98										٠				
	SEQUENC	CE L	ENGTI	H: 1	171	base	pai:	rs									
	SEQUENC																
15	STRANDE	EDNE	ss: c	doub	le												
	TOPOLOG	Y: :	linea	ar													
	ANTI-SE	ense	: No														
	ORIGINA	AL S	OURCE	3													
20	ORGANIS	5M: 1	Hepat	titis	3 C 1	virus	5										
20	IMMEDIA																
	CLONE:																
05	TG G	GG 2	ATC C	CCG 1	rat (	GAT A	ACC (	CGC 1	rgc :	rtt (	GAC S	rca :	ACG (	STC A	ACT (	GAG	47
25			le F														
		1				5			_		10					15	
	AAT	GAC	ATC	CGT	GTC	GAG	GAG	TCA	ATT	TAC	CAA	TGT	TGT	GAC	TTG		95
			Ile														
30					20		*			25		_	_	-	30		
	CCC	GAG	GCC	AGA	CAG	GCC	ATA	AGG	TCA	CTC	ACA	GAG	CGG	CTT	TAC	ATC	143
			Ala														
				35					40				_	45	-		
35	GGG	GGC	CCC	CTG	ACT	AAT	TCA	AAG	GGG	CAG	AAC	TGC	GGT	TAT	CGC	CGG	191
			Pro														
			50					<b>5</b> 5		*		-	60	-	_		
	TGC	CGC	GTC	AGC	GGC	GTG	CTG	ACG	ACT	AGC	TGC	GGT	AAT	ACC	CTC	ACA	239
40			Val														
		65					70				-	75					
	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	TGT	CGA	GCT	GCA	AAG	CTC	CAG	GAC	287
			TTG Leu														287

	TGC	ACG	ATG	CTT	GTG	TGC	GGA	GAC	GAC	CTT	GTC	GTT	ATC	arCar	САТ	AGC	335
	Cys	Thr	Met	Leu	Val	Cys	Gly	Asp	Asp	Leu	Val	Val	Ile	Cvs	Asn	Ser	555
					100			-	-	105				-1-	110		
5	GCG	GGA	ACT	CAG	GAG	GAC	GCG	GCG	AGC	CTA	CGA	GTC	TTC	ACG		GCT	383
	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	Ser	Leu	Arg	Val	Phe	Thr	Glu	Ala	
				115					120					125			
	ATG	ACT	AGG	TAC	TCT	GCC	CCC	CCC	GGG	GAC	CCG	CCC	CAA	CCA	GAA	TAC	431
10	Met	Thr	Arg	Tyr	Ser	Ala	Pro	Pro	Gly	Asp	Pro	Pro	Gln	Pro	Glu	Tyr	
			130					135					140				
												GTG					479
	Asp		Glu	Leu	Ile	Thr	Ser	Суз	Ser	Ser	Asn	Val	Ser	Val	Ala	His	
15		145					150					155	-				
												CGT					527
		Ala	Ser	Gly	Lys		Val	Tyr	Tyr	Leu	Thr	Arg	Asp	Pro	Thr	Thr	
	160					165					170					175	
20												CAC					575
	PIO	Leu	AIA	Arg		Ala	Trp	Glu	Thr		Arg	His	Thr	Pro	Val	Asn	
	ሞሮሮ	TO C	CTIA	ccc	180	<b>&gt;</b>	<b>&gt;</b> ma			185					190		
	Ser	Tro	Len	Glar	AAC	TIO	ATC	ATG	TAC	GCG	CCC	ACC	TTA	TGG	GCA	AGG	623
25	501	1-p	шоц	195	Abii	116	TTE	Met	200	Ala	Pro	Thr	Leu		Ala	Arg	
	ATG	ATT	CTG		ACC	CAC	ጥጥረ	ሙጥር		አመሮ	cmm	CTA	ccc	205	<b>a.</b> a	<i>a</i>	c
	Met	Ile	Leu	Met	Thr	His	Phe	Phe	Ser	Tla	LOU	Leu	Ala	CAG	CAG	CAA	671
			210					215	001	116	Deu		220	GTII	GIU	GIN	
30	CTT	GAA	AAA	GCC	СТА	GAT	TGT		ATC	TAC	GGG	GCC	-	ጥልሮ	TICC	Amm	719
	Leu	Glu	Lys	Ala	Leu	Asp	Cys	Gln	Ile	Tvr	Glv	Ala	Thr	Tur	Ser	Tlo	113
		225				-	230			-4-	1	235		-1-	501	110	
	GAG	CCA	CTT	GAC	CTA	CCT	CAG	ATC	ATT	CAA	CGA	CTC	CAC	GGT	CTT	AGC	767
35												Leu					
	240					245					250			-		255	
	GCA	TTT.	TCA	CTC	CAT	AGT	TAC	TCT	CCA	GGT	GAG	ATC	AAT	AGG	GTG	GCT	815
	Ala																
40					260					265					270		
												CGA					863
	Ser	Cys	Leu	Arg	Lys	Leu	Gly	Val	Pro	Pro	Leu	Arg	Val	Trp	Arg	His	
				275		•			280					285			
45	CGG	GCC	AGA	AGC	GTC	CGC	GCT	AAG	CTA	CTG	TCC	CAG	GGG	GGG	AGG	GCC	911

	Arg Ala Arg Ser Val Arg Ala Lys Leu Leu Ser Gln Gly Gly Arg Ala 290 295 300	
	GCC ACC TGT GGC AAA TAC CTC TTC AAC TGG GCA GTA AAG ACC AAG CTC 95	٥
5	Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Lys Thr Lys Leu	,
	305 310 315	
	AAA CTC ACT CCA ATC CCA GAA GCG TCC CAG CTG GAC TTG TCC GGC TGG 100	7
	Lys Leu Thr Pro Ile Pro Glu Ala Ser Gln Leu Asp Leu Ser Gly Trp	
10	320 325 330 335	
	TTC GTT GCT GGT TAC AGC GGG GGA GAC ATA TAT CAC AGC CTG TCT CGT 105	5
	Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg	
	340 345 350	
15	GCC CGA CCC CGC TGG TTC ATG TGG TGC CTA CTC CTA CTT TCC GTA GGG 110	3
	Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 355 360 365	
	355 360 365 GTA GGC ATC TAC CTG CTC CCC AAC CGA TGA GCG GGG AGC TAA ACA CTC 115	_
	Val Gly Ile Tyr Leu Leu Pro Asn Arg StopAla Gly Ser StopThr Leu	1
20	370 375 380	
	CAG GCC AAT AGG CCA TCC C C 117:	1
	Gln Ala Asn Arg Pro Ser	L
25	385	
	SEQ ID NO:99	
	SEQUENCE LENGTH: 1170 base pairs	
	SEQUENCE TYPE: nucleic acid	
30	STRANDEDNESS: double	
	TOPOLOGY: linear	
	ANTI-SENSE: NO	
	ORIGINAL SOURCE	
35	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
	CLONE: 030-2	
40	G GGG ATC CCG TAT GAT ACC CGC TGC TTT GAC TCA ACG GTC ACT GAG 46	5
	Gly Ile Pro Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
	1 5 10 15	
	AAT GAC ATC CGT GTT GAG GAG TCA ATT TAC CAA TGT TGT GAC TTG GCC 94	ŀ
45	Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
50		

					20					25			•		30		
	CCC	GAG	GCC	AGA	CAG	GCC	АТА	AGG	TCG		ACA	GAG	ccc	C Մ		ልጥሮ	142
						Ala											142
5				35				-	40				••••	45	-1-	110	
	GGG	GGC	CCC	CTG	ACT	AAT	TCA	AAA		CAG	AAC	TGC	GGC		CGC	CGG	190
						Asn											
	•		50					55	_				60	-4	5	5	
10	TGC	CGC	GTC	AGC	GGC	GTG	CTG	ACG	ACT	AGC	TGC	GGC	AAT	ACC	CTC	ACA	238
						Val											
		65					70				_	75					
	TGT	TAC	TTG	AAG	GCC	TCT	GCA	GCC	TGT	CGA	GCT	GCA	AAG	CTC	CAG	GAC	286
15	Cys	Tyr	Leu	Lys	Ala	Ser	Ala	Ala	Сув	Arg	Ala	Ala	Lys	Leu	Gln	Asp	
	80					85					90					95	
	TGC	ACG	ATG	CTT	GTG	TGC	GGA	GAC	GAC	CTT	GTC	GTT	ATC	TGT	GAA	AGC	334
	Cys	Thr	Met	Leu	Val	Cys	Gly	Asp	Asp	Leu	Val	Val	Ile	Суз	Glu	Ser	
20					100					105					110		
						GAC											382
	Ala	Gly	Thr	Gln	Glu	Asp	Ala	Ala	Ser	Leu	Arg	Val	Phe	Thr	Glu	Ala	
				115					120					125			
25						GCC											430
	Met	Thr		Tyr	Ser	Ala	Pro	Pro	Gly	qaA	Pro	Pro	Gln	Pro	Glu	Tyr	
	a.a		130					135					140				
						ACA											478
30	Asp		GIu	Leu	Ile	Thr		Сув	Ser	Ser	Asn	Val	Ser	Val	Ala	His	
	CNC	145	ma.	-			150			_		155					
						CGG											526
	160	VTG	Ser	GIĀ	гйа	Arg	vaı	ıyr	Tyr	Leu		Arg	Asp	Pro	Thr		
35		ርጣጥ	GCG	ccc	CCT	165	mcc	CNC	303	CCM	170	a. a	> am	~~~		175	
						GCG Ala											574
			2120	n g	180	VIG	тър	GIU	THE	185	Arg	nıs	Thr	PIO		Asn	
	TCC	TGG	СТА	GGC		ATC	ልጥሮ	ልጥር	ጥልጥ		ccc	NCC.	mm a	maa	190	200	caa
40						Ile											622
				195				1100	200	ma	110	1111	neu	205	ATG.	ALG	
	ATG	ATT	CTG		ACC	CAC	TTC	<b>ምም</b> ር		<b>ል</b> ሞሮ	Cum	CΦΔ	GCC		GNG	ממי	670
						His											370
45			210			<b>_</b>		215				u	220	3411	JIU	GIII	

	CTT	GAA	AAA	GCC	CTA	GAT	TGT	CAG	ATC	TAT	GGG	GCC	ACT	TAC	TCC	ATT	718
	Leu	Glu	Lys	Ala	Leu	Asp	Cys	Gln	Ile	Tyr	Gly	Ala	Thr	Tvr	Ser	Tle	
		225					230				_	235		-1-			
5	GAG	CCA	CTT	GAC	CTA	CCT	CAG	ATC	ATT	CAA	CGA	CTC	CAT	GGT	CTT	AGC	766
	Glu	Pro	Leu	Asp	Leu	Pro	Gln	Ile	Ile	Gln	Arg	Leu	His	Gly	Leu	Ser	
	240					245		•			250			-		255	
	GCA	TTT	TCA	CTC	CAT	AGT	TAC	TCT	CCA	GGT	GAG	ATC	AAT	AGG	GTG		814
10		Phe															
					260					265				- 3	270		
	TCA	TGC	CTC	AGG	AAA	CTT	GGG	GTA	CCG	CCC	TTG	CGA	GTC	TGG		САТ	862
	Ser	Cys	Leu	Arg	Lys	Leu	Gly	Val	Pro	Pro	Leu	Arg	Val	Trp	Arq	His	-
15				275					280					285			
	CGG	GCC	AGA	AGC	GTC	CGC	GCT	AAG	CTA	CTG	TCC	CAG	GGG	GGG	AGG	GCC	910
		Ala															
			290					295					300	-	-		
20	GCC	ACC	TGT	GGC	AAA	TAC	CTC	TTC	AAC	TGG	GCA	GTA	AAG	ACC	AAG	CTC	958
		Thr															
		305					310					315	_	•	•		
	AAA	CCC	ACT	CCA	ATC	CCG	GAA	GCG	TCC	CAG	CTG	GAC	TTG	TCC	GGC	TGG	1006
25		Pro															
	320					325					330				_	335	-
	TTC	GTT	GCT	GGT	TAC	AGC	GGG	GGA	GAC	ATA	TAT	CAC	AGC	CTG	тст	CGT	1054
		Val															
30					340					345					350		
	GCC	CGA	CCC	CGC	TGG	TTT	ATG	TGG	TGC	CTA	CTC	CTA	CTT	TCC	GTA	GGG	1102
	Ala	Arg	Pro	Arg	Trp	Phe	Met	Trp	Cys	Leu	Leu	Leu	Leu	Ser	Val	Gly	
				355					360					365			
<b>3</b> 5	GTA	GGC	ATC	TAC	CTG	CTC	CCC	AAC	CGA	TGA	GCG	GGG	AGC	TAA	ACA	CTC	1150
00	Val	Gly	Ile	Tyr	Leu	Leu	Pro	Asn	Arg	Stop	Ala	Gly	Ser	Stop	Thr	Leu	
			370					375					380				
	CAG	GCC	AAT	AGG	CCA	TCC	СС										1170
40	Gln	Ala	Asn	Arg	Pro	Ser											
40		385															

SEQ ID NO:100

SEQUENCE LENGTH: 1171 base pairs

45 SEQUENCE TYPE: nucleic acid

55

STRANDEDNESS: double

	TOPOL	OGY:	line	ear													
	ANTI-	SENSI	3: No	)													
5	ORIGI	NAL S	SOURC	CE													
	ORGAN	ISM:	Нера	atiti	s C	viru	ıs										
	IMMED:																
	CLONE																
10																	
	TG	GGG	ATC	CCG	TAT	GAT	ACC	CGC	TGC	TTT	GAC	TCA	ACA	GTC	АСТ	GAG	47
												Ser					• •
		1				5					10		•			15	
15	AA	GAC	ATC	CGI	GTT	GAG	GAG	TCA	ATT	TAC	CAA	TGT	TGT	GAC	TTG	GCC	95
												Cys					
					20					25					30		
	CCC	GAG	GCC	AGA	CAG	GCC	ATA	AGG	TCG	CTC	ACA	GAG	CGG	CTT	TAC	ATC	143
20	Pro	Glu	Ala	Arg	Gln	Ala	Ile	Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	
				35					40					45			
												TGC					191
	Gly	Gly	Pro	Leu	Thr	Asn	Ser	Lys	Gly	Gln	Asn	Cys	Gly	Tyr	Arg	Arg	
25			50					55					60				
	TGC	CGC	GCC	AGC	GGC	GTG	CTG	ACG	ACT	AGC	TGC	GGT	AAT	ACC	CTC	ACA	239
	Сув			Ser	Gly	Val			Thr	Ser	Сув	Gly	Asn	Thr	Leu	Thr	
		65					70					75					
30												GCA					287
			Leu	Lys	Ala		Ala	Ala	Сув	Arg	Ala	Ala	Lys	Leu	Gln	Asp	
	80 mca		3 ma			85					90					95	
												GTT					335
35	Cys	Thr	met	Leu		Сув	Gly	Asp	Asp		Val	Val	Ile	Сув	Glu	Ser	
	CCC	CCA	N CITT	030	100	~~~				105					110		
	Ala	Clar	Mb	CAG	GAG	GAC	GCG	GCG	AGC	CTA	CGA	GTC	TTC	ACG	GAG	GCT	383
	Ma	GIY	THE	115	GIU	Asp	ATA	ATS		Leu	Arg	Val	Phe		Glu	Ala	
40	ልሞር	ልሮሞ	NCC.		m/cm	C00	200	000	120					125			
												CCC					431
		-411.	130	TAT	OCT	WIG	FEO		GTĀ	Asp	Pro	Pro		Pro	Glu	Tyr	
	GAC	ליטולב		CTC	מיים	AC P	mc »	135	maa	mcc		a==	140				
	~.ic	113	SAG	CIG	WIW	ACA	TCA	TGC	TCC	TCC	AAT	GTG	TCG	GTC	GCG	CAC	479

55

45

50

Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His

		145					150					155					
	GAC	GCA	TCA	GGC	AAA	CGG	GTG	TAC	TAT	CTC	ACC			ccc	ccc	ACC	527
						Arg											
5	160					165		_			170		•			175	
	CCC	CTT	GCG	CGG	GCT	GCG	TGG	GAG	ACA	GCT	AGA	CAC	ACT	CCA	GTC		
						Ala											
					180					185					190		
10	TCC	TGG	CTA	GGC	AAC	ATC	ATC	ATG	TAC	GCG	CCC	ACC	TTA	TGG	GCA	AGG	623
						Ile											
				195					200					205			
	ATG	ATT	CTG	ATG	ACC	CAC	TTC	TTC	TCC	ATC	CTT	CTA	GCC	CAG	GAG	CAA	671
15	Met	Ile	Leu	Met	Thr	His	Phe	Phe	Ser	Ile	Leu	Leu	Ala	Gln	Glu	Gln	
			210					215					220				
						GAT											
	Leu		Lys	Ala	Leu	Asp	Сув	Gln	Ile	Tyr	Gly	Ala	Thr	Tyr	Ser	Ile	
20		225					230	•				235					
						CCT											767
	240	Pro	ren	Asp	Leu	Pro	Gln	Ile	Ile	Gln		Leu	His	Gly	Leu	Ser	
			ma.			245					250					255	
25						AGT											815
	nia	rne	ser	теп	260	Ser	ıyr	Ser	Pro		Glu	Ile	Asn	Arg		Ala	
	TCA	ጥርር	כיזיכי	AGC.		CTT	ccc	Cms	000	265	mma	~~~			270		
						Leu											863
30		-1-		275	270	Deu	GIY	Val	280	PIO	ren	Arg	Val		Arg	His	
	CGG	GCC	AGA		GTC	CGC	GCT	AAG		ርሞር	ሞሮሮ	CAC	ccc	285	200	000	011
						Arg											911
	_		290			- 3		295			001	0211	300	GLY	мy	MIG	
35	GCC	ACC	TGT	GGC	AAA	TAC	CTC	TTC	AAC	TGG	GCA	GTA		ACC	AAG	ርጥር	959
						Tyr											,,,
		305			_	_	310	•		•		315			-1-		•
	AAA	CTC	ACT	CCA	ATC	CCG	GAA	GCG	TCC	CAG	CTG	GAC	TTG	TCC	GGC	TGG	1007
40	Lys																
	320					325					330	_			-	335	
	TTC	GTT	GCT	GGT	TAC	AGC	GGG	GGA	GAC	ATA	TAT	CAC	AGC	CTG	тст		1055
	Phe																
<b>4</b> 5					340					345					350	-	

	GCC CGA CCC CGC TGG TTC ATG TGG TGC CTA CTC CTA CTT TCC GTA GGG 110	3
	Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly	-
	355 360 365	
5	GTA GGC ATC TAC CTG CTC CCC AAC CGA TGA GCG GGG AGC TAA ACA CTC 115	ı
	Val Gly Ile Tyr Leu Leu Pro Asn Arg StopAla Gly Ser StopThr Leu	•
	370 375 380	
	CAG GCC AAT AGG CCA TCC C C	1
10	Gln Ala Asn Arg Pro Ser	•
	385	
	SEQ ID NO:101	
15	SEQUENCE LENGTH: 7911 base pairs	
15	SEQUENCE TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
00	ANTI-SENSE: No	
20	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
	IMMEDIATE EXPERIMENTAL SOURCE	
0.5	CLONE: T7N1-30	
25		
	ACTAGTTAAT ACGACTCACT ATAGGGTGCC AGCCCCCTGA TGGGGGCGAC ACTCCACCAT 60	
	AGATCACTCC CCTGTGAGGA ACTACTGTCT TCACGCAGAA AGCGTCTAGC CATGGCGTTA 12	
00	GTATGAGTGT CGTGCAGCCT CCAGGACCCC CCCTCCCGGG AGAGCCATAG TGGTCTGCGG 18	
30	AACCGGTGAG TACACCGGAA TTGCCAGGAC GACCGGGTCC TTTCTTGGAT CAACCCGCTC 24	
	AATGCCTGGA GATTTGGGCG TGCCCCCGCG AGACTGCTAG CCGAGTAGTG TTGGGTCGCG 30	
	AAAGGCCTTG TGGTACTGCC TGATAGGGTG CTTGCGAGTG CCCCGGGAGG TCTCGTAGAC 36	0
05	CGTGCATC ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA ATC AAA CGT AAC 41	0
35	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Ile Lys Arg Asn	
	1 5 10	
-	ACC AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC GGT GGT CAG ATC 45	8
	Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile	
40	15 20 25 30	
	GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG TTG GGT GTG 50	6
	Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val	
	35 40 45	
45	CGC GCG ACT AGG AAG ACT TCC GAG CGG CCG CAA CCT CGT GGA AGG CGA 55	4

				50	ı				55			Pro		60			
	CAA	CCT	ATC	ccc	AAG	GCT	CGC	CAA	ccc	GAG	GGT	AGG	GCC	TGG	GCT	CAG	602
5	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Gln	Pro	Glu	Gly	Arg	Ala	Trp	Ala	Gln	
			65					70					75				
	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TTG	GGG	TGG	GCA	650
	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	Gly	Trp	Ala	
10		80					85					90					
	GGA	TGG	CTC	CTG	TCA	ccc	CGC	GGC	TCC	CGG	CCT	AGT	TGG	GGC	CCC	ACG	698
	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	Thr	
	95					100					105					110	
15	GAC	CCC	CGG	CGT	AGG	TCG	CGT	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	746
	Asp	Pro	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Авр	Thr	Leu	
					115					120					125		
	ACA	TGC	GGC	TTC	GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	794
20	Thr	Сув	Gly		Ala	qaA	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val	Gly	Ala	
				130					135					140			
	CCC	CTA	GGG	GGC	GCT	GCC	AGG	GCT	CTA	GCG	CAT	GGC	GTC	CGG	GTT	CTG	842
	Pro	Leu		Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	Gly	Val	Arg	Val	Leu	
25			145	_				150					155				
	GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAT	CTG	CCT	GGT	TGC	TCC	TTT	890
	GIU	Asp	GIY	Val	Asn	Tyr		Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	
		160					165					170					
30	TCT	ATC	TTC	CTT	TTG	GCT	TTG	CTG	TCC	TGT	TTG	ACC	ATC	CCA	GCT	TCC	938
	175	ше	Pne	ren	Leu		Leu	Leu	Ser	Сув	Leu	Thr	Ile	Pro	Ala	Ser	
	_	mag	<b>a</b>	oma.		180					185					190	
	31-	TAC	CAA	GIG	CGC	AAC	GCG	TCC	GGG	GTG	TAC	CAT	GTC	ACG	AAC	GAC	986
35	A10	тАт	GIII	vai		Asn	ATA	Ser	GLY		Tyr	His	Val	Thr	Asn	Asp	
	ምርር	መሮሮ	N N C	mas.	195	3 mm	ama			200					205		
	Cve	Sor	AAC	Com	AGT	ATT	GTG	TAT	GAG	GCG	GCG	GAC	GTG	ATT	ATG	CAC	1034
	Cyo	Ser	ASII	210	Ser	iie	vaı	Tyr		Ala	Ala	Asp	Val	Ile	Met	His	
40	ACC	ccc	ccc		CIMC.	000	maa	-	215					220		•	
	Thr	Pro	Clar	Crra	U-1	Doo	TGC	GTC	CGG	GAG	AAC	AAT	TCC	TCC	CGC	TGC	1082
	~***	-10	225	Cyb	AGI	FTO	CÀR		Arg	GIU	Asn	Asn		Ser	Arg	Cys	
	ጥርር	СПУ		CIDA	a Cm		200	230	000				235				
45	TOO	Va 1	31 <u>-</u>	LOU	mb-	Dro	Mb	CTT	GCG	GCC	AGG	AAC	AGC -	AGC	ATC	CCC	1130
		141	via	ren	THE	LT.O	THE	ren	А1а	Ala	Arg	Asn	Ser	Ser	Ile	Pro	

		240					245					250					
	ACT	ACG	ACA	ATA	CGG	CGT	CAT	GTC	GAC	TTG	CTC		GGG	GCA	CCT	CCT	1178
					Arg												1170
5	255				-	260					265		0-1			270	
	CTC	TGT	TCC	GCT	ATG	TAT	GTG	GGG	GAT	TTT	TGC	GGA	тст	GTT	ጥጥር		1226
					Met												
	٠				275	_		-	-	280	•				285		
10	GTC	TCC	CAG	CTG	TTC	ACT	TTC	TCA	ССТ	CGC	CGG	TAT	GAG	ACG	GTG	CAA	1274
					Phe												
				290					295			_	•	300			
	GAC	TGC	AAT	TGC	TCA	ATC	TAT	CCC	GGC	CAT	GTA	TCA	GGC	CAT	CGC	ATG	1322
15	Asp	Cys	Asn	Сув	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg	Met	
			305					310					315				
	GCT	TGG	GAT	ATG	ATA	ATG	AAT	TGG	TCA	CCT	ACA	ACA	GCC	CTA	GTG	GTA	1370
	Ala	Trp	Asp	Met	Ile	Met	Asn	Trp	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	
20		320					325					330					
					CGG												1418
	Ser	Gln	Leu	Leu	Arg	Ile	Pro	Gln	Ala	Val	Val	Asp	Met	Val	Ala	Gly	
	335					340					345					350	
25					GTC												1466
	Ala	His	Trp	Gly	Val	Leu	Ala	Gly	Leu		Tyr	Tyr	Ser	Met	Val	Gly	
		maa			355			_		360					365		
					GTC												1514
30	ASII	тър	AIA	ьув 370	Val	тел	vaı	vaı		Leu	Leu	Phe	Ala	_	Val	Asp	
	ccc	ccc	እሮሮ		CITIC	202	ccc	ccc	375	CM3	000	m> 0		380			
					GTG Val												1562
	011	017	385	1120	Val	Till	GIY	390	пув	vai	ATG	TYL		Thr	GIN	GIY	
35	TTT	ACA		<b>TTC</b>	TTT	ጥሮል	CGA		ccc	ጥርጥ	CAG	מממ	395	CAA	cam	CMN	1610
					Phe												1610
		400			0	001	405	G ₁	110	Der	GIII	410	116	GIII	rea	Val	
	AAC		AAC	GGC	AGC	TGG		ATC	ААТ	AGG	ACT		CTC	<b>ከ</b> ል መ	ጥርር	<b>ከልጥ</b>	1658
40					Ser												1030
	415			-		420				5	425				0,10	430	
	GAC	TCC	CTT	AAC	ACC	GGG	TTC	CTT	GCC	GCG		TTC	TAC	ACC	CAC		1706
					Thr												
45					435					440			•		445		

	TTC	AAC	GCG	TCC	GGA	TGT	CCG	GAG	CGT	ATG	GCC	GGT	TGC	CGC	CCC	ATT	1754
	Phe	Asn	Ala	Ser	Gly	Cys	Pro	Glu	Arg	Met	Ala	Gly	Сув	Arg	Pro	Ile	
				450					455					460			
5	GAC	GAG	TTC	GCT	CAG	GGG	TGG	GGT	CCC	ATC	ACT	CAT	GTT	GTG	CCT	AAC	1802
	Asp	Glu	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	Thr	His	Val	Val	Pro	Asn	
			465					470					475				
	ATC	TCG	GAC	CAG	AGG	CCC	TAT	TGC	TGG	CAC	TAC	GCG	CCT	CGA	CCG	TGT	1850
10	Ile	Ser	Asp	Gln	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Ala	Pro	Arg	Pro	Сув	
		480					485					490					
		ATC															1898
	Gly	Ile	Val	Pro	Ala	Ser	Gln	Val	Cys	Gly	Pro	Val	Tyr	Сув	Phe	Thr	
15	495					500					505					510	
		AGC															1946
	Pro	Ser	Pro	Val	Val	Val	Gly	Thr	Thr	qaA	Arg	Phe	Gly	Ala	Pro	Thr	
					515					520					525		
20		AAC															1994
	Tyr	Asn	Trp		Asn	Asn	Glu	Thr	Asp	Val	Leu	Leu	Leu	Asn	Asn	Thr	
				530					535					540			
		CCG															2042
25	Arg	Pro		Gln	Gly	Asn	Trp		Gly	Сув	Thr	Trp	Met	Asn	Gly	Thr	
			545					550					555				
		TTC															2090
	GIĀ	Phe	Thr	rĀs	Thr	Cys		Gly	Pro	Pro	Cys		Ile	Gly	Gly	Val	
30	000	560					565					570					
		AAC															2138
	575	Asn	ASII	THE	ren	580	cys	Pro	Thr	Asp		Pne	Arg	Lys	His		
		GCC	አ <i>ር</i> ሞ	መልሮ	אריא		mcm	CCM	maa	000	585	maa	mma			590	0106
35		Ala															2186
				-3-	595	nys	Сув	GIY	Ser	600	PIU	ттр	ren	THE	605	Arg	
	TGC	СТА	Cum	ጥልግ		CCA	<b>ጥ</b> ል ርግ	) ACC	CTC		CAC	ጠአጠ	ccc	maa		CMC	2224
		Leu															2234
40	-1-		•	610	-1-	110	-71	ALG	615	пр	птр	TÄT	PIO		THE	vaı	
	AAC	TTT	ACC		መጥር	AAG	COM	»CC		መአመ	CTC	ccc	ccc	620	CNN	CAC	2202
		Phe															2282
			625			-70	•41	630	ne t	T.Y.L	AGT	GTÄ	635	val	GIU	IITE	
45	AGG	СТТ		GCT	GCA	ጥርር	ልልጥ		<b>ACC</b>	CGA	CG3	GAG		መረንመ	GAC	നന്	2330
			~		3021	-00	·WI	100	ACC	CGA	JUA	UAU	CGI	IGI	GAC	116	233U

	Arg	Leu 640	Glu	Ala	Ala	Cys	Asn 645	Trp	Thr	Arg	Gly	Glu 650		Сув	Asp	Leu	
	GAG	GAC	AGG	GAT	AGA	TCA		Стт	AGC	cce	ርሞል			ሞሮሮ	ልሮአ	ACA	2378
5											Leu						2370
	655	_		-	•	660					665		200	501		670	
	GAG	TGG	CAG	GTA	CTG	CCC	TGT	TCC	TTC	ACC	ACC	CTG	CCG	GCT	CTG		2426
											Thr						2420
10					675		_			680					685	001	
	ACT	GGT	TTG	ATT	CAT	CTC	CAT	CAG	AAC	ATC	GTG	GAC	GTG	CAA		CTG	2474
											Val						
				690					695			-		700			
15	TAC	GGC	ATA	GGG	TCG	GCG	GTT	GTC	TCC	TTC	GCA	ATC	AAA	TGG	GAA	TAT	2522
											Ala						
			705					710					715	_		-	
	ATT	CTG	TTG	CTT	TTC	CTC	CCC	CTG	GCG	GAC	GCG	CGC	GTC	TGT	GCC	TGG	2570
20	Ile	Leu	Leu	Leu	Phe	Leu	Pro	Leu	Ala	Asp	Ala	Arg	Val	Сув	Ala	Trp	
		720					725					730					
	TTG	TGG	ATG	ATG	CTG	CTG	ATA	GCC	CAA	GCT	GAG	GCC	GCC	TTG	GAG	AAC	2618
	Leu	Trp	Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	Leu	Glu	Asn	
25	735			•		740					745					750	
20											GGA						2666
	Leu	Val	Val	Leu	Asn	Ala	Ala	Ser	Met	Ala	Gly	Ala	His	Gly	Ile	Leu	
					755					760					765		
30											TAC						2714
-	Ser	Phe	Leu		Phe	Phe	Сув	Ala	Ala	Trp	Tyr	Ile	Lys	Gly	Arg	Leu	•
				770					775					780			
											GTA						2762
35	val	Pro		Ala	Ala	Tyr	Ala		Tyr	Gly	Val	Trp	Pro	Leu	Leu	Leu	
•	CMC	mmo	785	000				790					795				
											GCC						2810
	rea		Met	Ala	Leu	Pro		Arg	Ala	Туг	Ala		Asp	Arg	Glu	Met	
40	CCT	800	mcc	mcc.	CC.	000	805	a				810					
											GGT						2858
	815	via	Ser	cys	αтĀ		ATG	vaı	rue	val	Gly	Leu	Val	Leu	Leu		
		ጥር አ	CCA	መልሮ	ma <i>c</i>	820	CITIC	mmc	ama	00=	825					830	
45											AAG						2906
<del>7</del> ∪	204	JUL	110	- A T	TÄT	пур	val	z iie	ren	ATG	Lys	тел	тте	Trp	Trp	Leu	

					835					840					845		
	CAA	TAT	CTC	ATC	ACC	AGG	GCC	GAG	GCG	CAC	TTG	CAA	GTG	TGG	ATC	CCC	2954
							Ala										
5				850					855					860			
	CCC	CTC	AAC	GTT	CGG	GGG	GGC	CGC	GAT	GCC	ATC	ATC	CTT	CTC	ACA	TGT	3002
							Gly										
	٠		865					870					875			-4	
10	GCG	GTC	CAC	CCG	GAG	CTG	ATC	TTT	GAC	ATC	ACC	AAG	CTC	TTG	CTC	GCC	3050
							Ile										
		880					885					890					
	ATA	CTC	GGT	CCG	CTC	ATG	GTA	CTC	CAG	GCT	GGC	CTA	ACC	CAA	ATG	CCG	3098
15							Val										
	895					900					905					910	
	TAC	TTT	GTG	CGT	GCT	CAA	GGG	CTC	ATT	CGT	ATG	TGC	ATG	TTG	GTG	CGG	3146
							Gly										
20					915					920					925	•	
	AAA	GCC	GCT	GGG	GGT	CAT	TAT	GTC	CAG	ATG	GCT	CTC	ATG	AAG	CTG	GCT	3194
	Lys	Ala	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Leu	Met	Lys	Leu	Ala	
				930					935					940			
25	GCA	CTG	ACA	GGT	ACG	TAC	GTT	TAT	GAC	CAT	CTT	ACT	CCA	CTG	CAG	GAC	3242
	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Asp	His	Leu	Thr	Pro	Leu	Gln	Asp	
			945					950					955				
	TGG	GCC	CAC	GCG	GGC	CTA	CGA	GAC	CTT	GCG	GTA	GCA	GTT	GAG	CCC	GTT	3290
30	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	Glu	Pro	Val	
		960					965					970					
							ACT										3338
	Val	Phe	Ser	Asp	Met	Glu	Thr	Lys	Ile	Ile	Thr	Trp	Gly	Ala	Glu	Thr	
35	975					980					985					990	
••	GCG	GCG	TGT	GGG	GAC	ATC	ATC	TCG	AGT	CTA	CCC	GTT	TCC	GCC	CGA	AGG	3386
	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Ser	Ser	Leu	Pro	Val	Ser	Ala	Arg	Arg	
					995					.000					.005		
40	GGG	AGG	GAG	CTG	CTT	TTG	GGA	CCG	GCC	GAT	AGT	TTT	GAC	GGG	CAG	GGG	3434
40	Gly	Arg	Glu	Leu	Leu	Leu	Gly	Pro	Ala	Asp	Ser	Phe	Asp	Gly	Gln	Gly	
				010					015					1020			
							ATC										3482
45	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ser	Gln	Gln	Thr	Arg	Gly	
<b>4</b> 5		1	.025				1	030				1	035				

	CTG	CTT	GGT	TGC	ATC	ATC	ACC	AGC	CTT	ACG	GGC	CGG	GAT	AAG	AAC	CAG	3530
																Gln	
		1040					1045					1050		-			
5	GTC	GAG	GGG	GAG	GTT	CAA	GTG	GTC	TCT	ACC	GCA	ACA	CAA	TCT	TTC	CTG	3578
					Val												
	1055					1060					1065					1070	
	GCG	ACC	TGC	ATC	AAC	GGC	GTT	TGC	TGG	ACT	GTT	TTC	CAC	GGC	GCC	GGC	3626
10	Ala	Thr	Cys	Ile	Asn	Gly	Val	Cys	Trp	Thr	Val	Phe	His	Gly	Ala	Gly	
					1075					1080					1085	_	
	TCG	AAG	ACC	TTA	GCC	GGC	CCA	AAA	GGC	CCA	ATC	ACC	CAA	ATG	TAC	ACC	3674
	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	Pro	Ile	Thr	Gln	Met	Tyr	Thr	
15				1090					1095					1100			
	AAT	GTA	GAT	CAG	GAC	CTC	GTC	GGC	TGG	TCG	GCG	CCC	CCC	GGG	GCG	CGT	3722
	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	Ser	Ala	Pro	Pro	Gly	Ala	Arg	
			1105					1110					1115				
20					TGC												3770
			Thr	Pro	Сув	Thr	Сув	Gly	Ser	Ser	Asp	Leu	Tyr	Leu	Val	Thr	
		L120					1125					1130					
					GTC												3818
25	Arg	His	Ala	Asp	Val	Ile	Pro	Val	His	Arg	Arg	Gly	Asp	Ser	Arg	Gly	
20	1135					1140					1145					1150	
					CCC												3866
	Ser	Leu	Leu	Ser	Pro	Gly	Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	
30					1155					1160					1165		
30					TGC												3914
	Gly	Pro			Сув	Pro	Ser	Gly	Arg	Val	Val	Gly	Ile	Phe	Arg	Ala	
				1170					175					1180			
					CGG												3962
35	Ala			Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	qaA	Phe	Val	Pro	Val	
			1185					1190					195				
					ACC												4010
			Met	Glu	Thr			Arg	Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	
<b>40</b>		200					205					210					
					GTA												4058
		Pro	Pro	Ala	Val		Gln	Thr	Phe			Ala	His	Leu	His	Ala	
	1215					.220					225					230	
45	CCC	ACT	GGC	AGC	GGC	AAA	AGC	ACC	AGG	GTG	CCG	ርርጥ	CCC	ጥልጥ	CCC	CCC	4106

	Pro	Thr	Gly				Ser	Thr	Arg	Val	Pro	Ala	Ala	Tyr	Ala	Ala	
					1235					1240					1245		
	CAA	GGG	TAC	AAG	GTA	CTC	GTC	CTG	AAC	CCG	TCC	GTT	GCT	GCC	ACT	TTG	4154
5	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	
				1250					1255					1260			
	GGC	TTT	GGG	GCG	TAC	ATG	TCC	AAG	GCA	CAT	GGT	GTT	GAC	CCT	AAC	ATC	4202
	Gly	Phe	Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Val	Asp	Pro	Asn	Ile	
10			1265					1270					1275				
	AGA	ACT	GGG	GTG	AGG	ACC	ATC	ACC	ACG	GGC	GCT	CCC	ATC	ACG	TAC	TCC	4250
	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	Pro	Ile	Thr	Tyr	Ser	
		1280					1285			•	:	1290					
15	ACC	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	GGT	GCC	TAT	4298
	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	
	1295				:	1300				:	1305					1310	
	GAC	ATC	ATA	ATA	TGT	GAT	GAG	TGT	CAT	TCA	ACT	GAC	TCG	ACT	TCC	ATC	4346
20	Asp	Ile	Ile	Ile	Сув	Asp	Glu	Сув	His	Ser	Thr	Asp	Ser	Thr	Ser	Ile	
					1315				:	1320					1325		
	TTG	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG	GAG	ACG	GCT	GGA	GCG	CGC	4394
	Leu	Gly	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	
25			:	1330				:	1335				:	1340			
25	CTT	GTC	GTG	CTC	GCC	ACC	GCT	ACG	CCT	CCG	GGA	TCG	GTC	ACC	GTG	CCG	4442
	Leu	Val	Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	Thr	Val	Pro	
		:	1345				;	1350				:	1355				
	CAT	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA	GAG	ATC	CCC	4490
30	His	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Asn	Thr	Gly	Glu	Ile	Pro	
	:	1360				:	1365				1	1370					
	TTC	TAT	GGC	AAG	GCC	ATC	CCC	CTC	GAG	GCC	ATC	AAG	GGG	GGG	AGG	CAT	4538
	Phe	Tyr	Gly	Lys	Ala	Ile	Pro	Leu	Glu	Ala	Ile	Lys.	Gly	Gly	Arg	His	
35	1375					1380				1	L385					1390	
	CTC	ATT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	GCG	AAG	4586
	Leu	Ile	Phe	Cys	His	Ser	Lys	Lys	Lys	Cy.s	Asp	Glu	Leu	Ala	Ala	Lys	
					1395				]	L <b>40</b> 0				:	1405		
40	CTG	TCG	GCC	CTC	GGA	GTC	AAC	GCT	GTA	GCA	TAT	TAC	CGG	GGT	CTT	GAT	4634
	Leu	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	
				1410					L <b>415</b>					L420		_	
	GTG	TCC	ATC	ATA	CCG	ACA	AGC	GGG	GAC	GTC	GTT	GTC	GTG	GCA	ACA	GAC	4682
45															Thr		

			1425					1430					1435				
	GCT	CTA	ATG	ACA	GGC	TAT	ACC	GGT	GAC	TTT	GAC	TCG	GTG	ATC	GAC	TGC	4730
	Ala	Leu	Met	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	
5		1440					1445					1450.			•	•	
	AAC	ACA	TGT	GTC	ACC	CAA	ACA	GTC	GAT	TTC	AGC	TTG	GAC	CCT	ACT	TTC	4778
		Thr															
	1455					1460					1465					1470	
10		ATC															4826
	Thr	Ile	Glu	Thr	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	
				-	1475					1480					1485		
		CGA															4874
15	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Gly	Gly	Ile	Tyr	Arg	Phe	Val	
				1490					1495					1500			
		CCA															4922
	unr	Pro		Glu	Arg	Pro			Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	
20	C) )		1505	<b>a</b> > a				1510					1515				
		TGT															4970
		Cys 1520	TYL	Asp	Ala			Ala	Trp	Tyr			Thr	Pro	Ala	Glu	
			രത്ത	NCC.	mmc		1525	m> a	<i>a</i> m>			1530					
25		TCG															5018
	1535	Ser	•441	my		540	ATG	TYL	ren			PIO	GIĀ	Leu			
		CAG	GAC	САТ			חיווירי	TICC.	CNC		1545 Cmc	mmc	3.00	000		1550	
		Gln															5066
30	-				.555	<b></b>				1560	Vai	FILE	IIII		1565	THE	
	CAC	ATA	GAT			TTC	TTG	TCC			444	CAG	GCA			አክሮ	5114
		Ile															2114
				1570					575					.580	1102		
35	TTC	CCC	TAC	CTG	GTA	GCA	TAC	CAG	GCT	ACA	GTG	TGC			GCC	AAG	5162
		Pro															
	*		585					590					595	-		•	
	GCT	CCA	CCT	CCA	TCG	TGG	GAT	CAG	ATG	TGG	AAG	TGT	CTC	ATA	CGG	CTG	5210
40		Pro															
	1	600				1	.605				1	610 ⁻					
		CCT															5258
		Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	
45	1615					620					625					630	

	GTT	CAG	AAC	GAG	GTT	ACC	CTC	ACA	CAC	CCC	ATA	ACC	AAG	TTC	ATC	ATG	5306
				Glu													
					1635					1640				-	1645		
5	GCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACT	TGG	GTG	CTG	5354
	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	
				1650					1655					1660			
				GTC													5402
10	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	Gly	
			1665					1670					1675				
				ATT													5450
			Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	Val	
15		1680					1685					1690					
				AGG													5498
		Pro	Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	
	1695		•			1700					1705					1710	
20				CAC													5546
	Cys	Ala	Ser	His		Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu	
	<b></b>	mm.a			1715					1720					1725		
				CAG													5594
25	GIN	Pne		Gln	Lys	Ala	Leu			Leu	Gln	Thr	Ala	Thr	Lys	Gln	
	CCC	CAC		1730	com		~		1735					L740			
				GCT													5642
	AIG		1745	Ala	Ala	Pro			GIu	Ser	Lys			Ala	Leu	Glu	
30	ACC			ccc	330	CNC		1750		-			1755				
				GCG													5690
		1760	110	Ala	цув		Met 1765	пр	ASR	Pne			GIY	TTE	GIn	Tyr	
			GGC	TTG	ምርር			CCM	CCA	አክሮ		770	a m a	CON	ma v	OMC.	<b></b>
35				Leu													5738
	1775		1			780	204		GLY		1785	ALG	116	ATG		1790	
	ATG	GCA	TTC	ACA			АТС	ACC	AGC			ልሮሮ	<b>ACC</b>	ሮአአ			5786
				Thr													3700
40					.795					1800					805	THE	•
	CTC	CTG	TTT	AAC	ATC	TTG	GGG	GGA			GCC	GCC	CAA			CCC	5834
				Asn													JVJ1
	٠			1810			-	-	1815					1820			
45	CCC	AGT	GCC	GCT	TCA	GCC	TTC	GTG	GGC	GCC	GGT	ATA	GCT			GCT	5882

	Pro	Ser	Ala	Ala	Ser	Ala	Phe	Val	Gly	Ala	Gly	Ile	Ala	Gly	Ala	Ala	
			1825					1830					1835				
	GTT	GGC	AGC	ATA	GGC	CTC	GGG	AAG	GTG	CTT	GTG	GAC	ATT	CTG	GCG	<b>G</b> GT	5930
5	Val	Gly	Ser	Ile	Gly	Leu	Gly	Lys	Val	Leu	Val	Asp	Ile	Leu	Ala	Gly	
		1840					1845					1850					
	TAT	GGA	GCA	GGG	GTG	GCA	GGC	GCG	CTC	GTG	GCC	TTT	AAG	GTC	ATG	AGC	5978
	Tyr	Gly	Ala	Gly	Val	Ala	Gly	Ala	Leu	Val	Ala	Phe	Lys	Val	Met	Ser	
10	1855					1860					1865					1870	
	GGT	GAC	ATG	CCC	TCC	ACC	GAG	GAC	CTG	GTC	AAC	TTA	CTC	CCC	GCC	ATC	6026
	${f Gly}$	Asp	Met	Pro	Ser	Thr	Glu	Asp	Leu	Val	Asn	Leu	Leu	Pro	Ala	Ile	
				;	1875				:	1880					1885		
15	CTC	TCT	CCT	GGT	GCC	CTG	GTC	GTC	GGG	GTC	GTG	TGC	GCA	GCA	ATA	CTG	6074
	Leu	Ser	Pro	Gly	Ala	Leu	Val	Val	Gly	Val	Val	Сув	Ala	Ala	Ile	Leu	
			:	1890				:	1895				;	1900			
	CGT	CGG	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG	TGG	ATG	AAC	CGG	6122
20	Arg	Arg	His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met	Asn	Arg	
		:	1905					1910				:	1915				
	CTG	ATA	GCG	TTT	GCT	TCG	CGG	GGC	AAC	CAT	GTC	TCC	CCC	ACG	CAC	TAT	6170
	Leu	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val	Ser	Pro	Thr	His	Tyr	
25	:	1920				;	1925				1	1930					
	GTG	CCT	GAA	AGC	GAC	GCC	GCA	GCG	CGC	GTC	ACC	CAG	ATC	CTC	TCC	AAC	6218
	Val	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	Asn	
	1935				1	1940				1	1945				:	1950	
30	CTT	ACC	ATC	ACT	CAG	CTG	TTG	AAG	AGG	CTT	CAC	CAG	TGG	ATT	AAT	GAG	6266
00	Leu	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	Gln	Trp	Ile	Asn	Glu	
				1	1955				1	1960				2	1965		
	GAC	TGC	TCC	ACG	CCA	TGC	TCC	GGC	TCG	TGG	CTC	AGG	GAT	GTT	TGG	GAC	6314
35	Asp	Сув	Ser	Thr	Pro	Cys	Ser	${\tt Gly}$	Ser	$\mathbf{Trp}$	Leu	Arg	Asp	Val	Trp	Asp	
33			:	1970				1	1975				1	1980			
	TGG	ATA	TGC	ACG	GTA	TTG	GCT	GAT	TTC	AAG	ACC	TGG	CTC	CAG	TCC	AAG	6362
	Trp	Ile	Сув	Thr	Val	Leu	Ala	Asp	Phe	Lys	Thr	Trp	Leu	Gln	Ser	Lys	
		1	1985				:	1990				1	1995				
40	CTC	CTG	CCG	CGG	TTA	CCG	GGG	GTC	CCT	TTT	TTC	TCA	TGC	CAG	CGT	GGG	6410
	Leu	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	Ser	Сув	Gln	Arg	Gly	
	:	2000				:	2005	•			2	2010					
	TAC	AAG	GGG	GTT	TGG	CGG	GGA	GAT	GGC	ATC	ATG	TAT	ACC	ACC	TGC	CCA	6458
45	Tyr	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	Tyr	Thr	Thr	Cvs	Pro	

	2015				•	2020					2025					2030	
	TGT	GGA	GCA	CAA	ATC	ACC	GGA	CAT	GTC	AAA	AAC	GGT	тст	ATG			6506
5														Met			
Ū					2035					204.0		-			2045		
	GTT	GGG	ССТ	AGA	ACC	TGT	AGC	AAC	ACG	TGG	CAC	GGA	ACA	TTT	CCC	ATC	6554
														Phe			
10	•			2050					2055	_		-		2060			
10	AAC	GCG	TAC	ACC	ACA	GGC	CCC	TGC	ACA	CCC	TCC	CCG	GCG	CCA	AAC	TAT	6602
														Pro			
			2065					2070					2075			•	
	TCC	AGG	GCG	TTG	TGG	CGG	GTG	GCC	GCT	GAG	GAG	TAT	GTG	GAG	GTC	ACG	6650
15	Ser	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ala	Glu	Glu	Tyr	Val	Glu	Val	Thr	•
	:	2080				:	2085				:	2090					
	CGG	GTG	GGG	GAT	TTC	CAC	TAC	GTG	ACG	GGC	ATG	ACC	ACT	GAC	AAC	GTG	6698
••	Arg	Val	Gly	Asp	Phe	His	Tyr	Val	Thr	Gly	Met	Thr	Thr	Asp	Asn	Val	
20	2095					2100					2105					2110	
														GAA			6746
	Lys	Сув	Pro	Сув	Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	
25					2115					2120					2125		
25														CTC			6794
	Gly	Val			His	Arg	Tyr	Ala	Pro	Ala	Сув	Lys	Pro	Leu	Leu	Arg	
				2130					2135					2140			
														GTT			6842
30	Asp			Thr	Phe	Gln			Leu	Asn	Gln	Tyr	Thr	Val	Gly	Ser	
	a.a		2145					2150					2155				
														ACC			6890
			Pro	Сув	GIU			Pro	Asp	Val			Val	Thr	Ser	Met	
35		2160	CNC	000	maa		2165				_	2170					
														CGT			6938
•	2175	TILL	Asp	PIO			TTE	THE	ATA			ALA	Arg	Arg			
	_	202	ccc	mcm.	-	2180	maa	mmo	000		2185					2190	
40														AGT			6986
	wid	wrd	GTĀ			PTO	ser	ren			ser	ser	Ala	Ser		Leu	
	ጥርጥ	CCC	Cutt		2195	ma.c	ccc	3.03		2200	100	<b>~~</b>	<b>~~</b>		2205		<b>5024</b>
														GGC			7034
45	Set	vra			Ten	scop	PTE			Thr	TUL	H18 [.]		Gly	Ala	Pro	
			4	2210				2	215				2	2220			

	GAC	ACT	GAC	CTC	ATC	GAG	GCC	AAC	CTC	CTG	TGG	CGG	CAG	GAG	ATG	GGC	7082
	Asp	Thr	Asp	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	
			2225					2230					2235				
5								TCA									7130
	Gly	Asn	Ile	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	Ile	Val	Ile	Leu	Asp	
		2240					2245					2250					
								GAG									7178
10	Ser	Phe	Glu	Pro	Leu	Arg	Ala	Glu	Glu	Asp	Glu	Arg	Glu	Val	Ser	Val	
	2255					2260					2265					2270	
								ACC									7226
	Ala	Ala	Glu	Ile	Leu	Arg	Lys	Thr	Arg	Lys	Phe	Pro	Ala	Ala	Met	Pro	
15					2275					2280					2285		
								AAC									7274
	Val	Trp	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	Leu	Glu	Ser	Trp	Lys	
				229	-				229	_				230	-		
20								GTG									7322
	Asn			Tyr	Val	Pro		Val	Val	His	Gly	Сув	Pro	Leu	Pro	Pro	
			2305					2310					2315				
								CCT									7370
25			Ala	Pro	Pro			Pro	Pro	Arg			Arg	Thr	Val	Val	
		2320					2325					2330					
								TCT									7418
		Thr	GIu	Ser			Ser	Ser	Ala			Glu	Leu	Ala		_	
30	2335	mmm	000	`		2340					2345					2350	
								TCG									7466
	THE	Pne	GTĀ			GIY	Ser	Ser			Asp	Ser	Gly			Thr	
	ccc	CCM	CCM		2355	G00	maa	000		2360					2365		
35								GCC									7514
	GIY	PIO		Asp 2370	GIN	Ala	ser	Ala		GIY	Asp	Ala			Asp	Ala	
	GAG	ምሮር			mcc.	እመሮ	000		2375	G2.G	223	~~~		2380			
								CCC									7562
40	Giu		2385	Ser	Ser	Met		Pro	ren	GIU	GIY			GIY	Авр	Pro	
	GAT			GAC	GGG	ጥርጥ		2390 TCT	NCC.	Cmx	אככ		2395	000	200	020	76.0
								Ser									7610
		2400		-10P	GLY		405	SAT	TIIL	AGI		G1u 2410	GIU	WTG	ser	GTU	
45			GTC	TGC	TGC			ጥሮሮ	ጥልሮ	ልሮል			ccc	ccc	mm s	እ mm	7650

			Val	Cys				Ser	Tyr	Thr	Trp	Thr	Gly	Ala	Leu	Ile	
	2415					2420					2425					2430	
5																AGC	7706
3	Thr	Pro	Суѕ			GLu	Glu	Ser			Pro	Ile	Asn	Ala	Leu	Ser	
	***	aam			2435					2440					2445		
																CGC	7754
	Asn	Pro	Leu		Arg	His	His			Val	Tyr	Ala	Thr	Thr	Ser	Arg	
10				2450					2455					2460			
			AGC														7802
	Ser		Ser	Gln	Arg	Gln			Val	Thr	Phe			Leu	Gln	Val	
	~~~		2465					2470					2475				
15			GAC														7850
•			Asp	His	Tyr			Val	Leu	Lys	Asp	Met	Lys	Ala	Lys	Ala	
		2480					2485					2490					
			GTT														7898
20		Thr	Val	Lys			Leu	Leu	Ser			Glu	Ala	Сув	Lys	Leu	
	2495					2500				2	2505				:	2510	
	ACG	ccc	CCA	CAC	T	7911											
25	SEQ ID SEQUENC SEQUENC STRANDE	CE LI	ENGTI YPE:	nuc	leic		_	rs									
30	TOPOLOG ANTI-SE ORIGINA	GY: : BNSE: AL SO	linea : No DURCE	ar S													
35	ORGANIS IMMEDIA CLONE:	ATE I	EXPER														
40	AAGO		ATG A Met S								urg I						48
	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC			GGT	GCT	CAG	ል ሞሮ	96
			Arg														30 ,
	15			. 3	, _	20	P		-1-		25	1	1	~~ <i>I</i>		30	
45		GGT	GGA	GTT	TAC		ጥ ፕር	CCG	CGC	AGG		ccc	AGG	արարաշ	CCT		144
													0		331	316	744

5

	Val	Gly	Gly	Val		Leu	Leu	Pro	Arg			Pro	Arg	Leu	Gly	Val	
	ccc	ccc	» cm	100	35	3.00				40					45		
_	y x c	υ GCG Nla	Mb~	AGG	AAG	ACT	TCC	GAG	CGG	TCG	CAA	CCT	CGT	GGA	AGG	CGA	192
5	ALG	Ala	THE	50	гЛя	Thr	ser	GIU	Arg 55	Ser	GIn	Pro	Arg	Gly 60	Arg	Arg	
	CAA	ССТ	ATC		AAG	GCT	CGC	CAA		GAG	ccc	ACC.	ccc		ccm	CNC	240
		Pro															240
10			65				9	70		024	GLY	ary	75	тър	ALG	GIII	
70	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	_	GGG	TGG	GCA	288
		Gly															200
		80					85	_	_			90.					
15	GGA	TGG	CTC	CTG	TCA	CCC	CGG	GGG	GTT	GCG	AAG	GCG	GTG	GAC	TTT	GTG	336
	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Val	
	95					100					105					110	
		GTT															384
20	Pro	Val	Glu	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Thr	Asp	
					115					120					125		
		TCA															432
	Asn	Ser	Thr		Pro	Ala	Val	Pro		Thr	Phe	Gln	Val	Ala	His	Leu	
25	010	aam	000	130					135					140			
		GCT															480
	urs	Ala	145	Thr	GIY	ser	Gly		Ser	Thr	Arg	Val		Ala	Ala	Tyr	
	GCG	GCC	-	ecc	መእሮ	አክሮ	CMX	150	ama	ama.		000	155				
30		Ala															528
		160		G_1	-1-	Dy 6	165	neu	Vai	red	ASII	170	Ser	vaı	ALA	Ala	
	ACT	TTG	GGC	TTT	GGG	GCG		АТС	ጥርር	AAG	GCA		CCT	cmm	CNC	CCm	576
		Leu															376
35	175		-		•	180				_1_	185		01	TU	rwp	190	
	AAC	ATC	AGA	ACT	GGG	GTG	AGG	ACC	ATC	ACC		GGC	GCT	CCC	ATC		624
		Ile															021
					195					200		-			205		
40	TAC	TCC	ACC	TAC	GGT	AAG	TTC	CTC	GCC	GAC	GGT	GGC	TGT	TCT	GGG	GGT	672
		Ser															
				210					215					220	_	٠	
		TAT															720
45	Ala	Tyr	Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	Asp	Ser	Thr	

			225					230					235				
	TCC	ATC	TTG	GGC	ATT	GGT	ACA	GTC	CTG	GAC	CAA	GCG		ACG	GCT	GGA	768
							Thr										
5		240					245			-		250				1	
	GCG	CGC	CTT	GTC	GTG	CTC	GCC	ACC	GCT	ACG	ССТ	CCG	GGA	TCG	GTC	ACC	816
							Ala										
	255					260					265		-			270	
10	GTG	CCG	CAT	CCT	AAT	ATT	GAG	GAG	GTG	GCC	TTG	TCC	AAC	ACT	GGA	GAG	864
							Glu										
•					275					280					285		
	ATC	CCC	TTC	TAT	GGC	AAG	GCC	ATC	CCC	CTC	GAG	GCC	ATC	AAG	GGG	GGG	912
15	Ile	Pro	Phe	Tyr	Gly	Lys	Ala	Ile	Pro	Leu	Glu	Ala	Ile	Lys	Gly	Gly	
				290					295					300			
	AGG	CAT	CTC	ATT	TTC	TGC	CAT	TCC	AAG	AAG	AAA	TGT	GAC	GAG	CTC	GCT	960
	Arg	His	Leu	Ile	Phe	Cys	His	Ser	Lys	Lys	Lys	Сув	Asp	Glu	Leu	Ala	
20			305					310					315				
							GGA										1008
	Ala	Lys	Leu	Ser	Ala	Leu	Gly	Val	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	
		320					325					330		•			
25							CCG										1056
		Asp	Val	Ser	Ile	Ile	Pro	Thr	Ser	Gly	Asp	Val	Val	Val	Val	Ala	
	335					340					345					350	
							GGC										1104
30	Thr	Asp	Ala	Leu		Thr	Gly	Tyr	Thr	Gly	qaA	Phe	Asp	Ser	Val	Ile	
					355					360					365		
							AGAT	CT									1128
	Asp	Cys	Asn		Stop	Stop	•										
35				370													

SEQ ID NO:103

SEQUENCE LENGTH: 974 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

ANTI-SENSE: No

ORGANISM: Hepatitis C virus
IMMEDIATE EXPERIMENTAL SOURCE

50

CLONE: 015-1

	GCG	GATC	CT C	CA C	СТ С	CA T	CG T	GG G	AC C	AA A	TG T	GG A	ag T	GT C	TC A	TA C	GG 51
5																	
			P		ro P	ro S	er T	rp A	sp G	ln M	et T	rp L	ys C	ys L	eu I	le A	rg
				1				5					10				
			CCT														99
10			Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	
	15					20					25					30	
			CAA														147
	Ala	Val	Gln	Asn	Glu	Val	Thr	Leu	Thr	His	Pro	Ile	Thr	Lys	Phe	Ile	
15					35					40		•			45		
	ATG	GCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACC	TGG	GTG	195
	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	
				50					55					60			
20			GGC														243
	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Сув	Leu	Thr	Thr	
			65					70					75				
			GTG														291
25	Gly	Ser	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Arg	Pro	Ala	
		80					85					90					
			CCC														339
	Ile	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Arg	Glu	Phe	Asp	Glu	Met	Glu	
30	95					100					105					110	
••			GCC														387
	Glu	Cys	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	
					115					120					125		
35			TTC														435
J J	Glu	Gln	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr	Ala	Thr	Gln	
				130					135			,		140			
			GAG										-				483
	Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	Trp	Arg	Ala	Leu	
40			145					150					155				
			TTC														531
	Glu	Ala	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	
		160					165					170					
45	TAC	TTG	GCA	GGC	TTG	TCC	ACT	CTG	CCT	GGA	AAC	CCC	GCG	ATA	GCA	TCA	579

50

		Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly		Pro	Ala	Ile	Ala	Ser	
	175					180					185					190	
_						GCC											627
5	rea	Mec	ATG	Pne		Ala	Ser	He	Thr		Pro	Leu	Thr	Thr		His	
	ልሮሮ	CTC	ሮሞር	trono	195	N III C	mmc	000	CCN	200	Om.	000			205		
						ATC Ile											675
10			Dou	210	111)11	116	neu	GIX	215	ттр	vai	AIG	ATG	220	ren	AIA	
70	CCT	CCC	AGC		GCT	TCA	GCT	անար		GGC	GCC	ccc	מיים		ccc	ccc	772
						Ser											123
			225					230		1		0- 1	235		GLY	nia	
15	GCT	GTT	GGC	AGC	ATA	GGC	CTT	GGG	AAG	GTG	CTT	GTG		ATC	CTG	GCG	771
						Gly											
		240					245					250					
	GGT	TAT	GGA	GCA	GGG	GTG	GCA	GGC	GCA	CTC	GTG	GCC	TTT	AAG	GTC	ATG	819
20	Gly	Tyr	Gly	Ala	Gly	Val	Ala	Gly	Ala	Leu	Val	Ala	Phe	Lys	Val	Met	
	255					260					265		•			270	
						TCC											867
	Ser	Gly	Glu	Met		Ser	Thr	Glu	Asp	Leu	Val	Asn	Leu	Leu	Pro	Ala	
25					275					280					285		
						GCC											915
	116	ren	ser		GIA	Ala	Leu	Val		Gly	Val	Val	Сув		Ala	Ile	
	CTC	CCm	CCA	290	cmc	ccc	003	<i>a</i> aa	295					300			
30						GGC											963
	Dea	my	305	HIS	AGI	Gly	PIO	310	GIU	GIA	AIA	vaı	315	Trp	Met	Asn	
	CGG	CTG	C AC	CC				310					313				974
	Arg																314
35		320															
	SEQ ID	NO: 1	104														
	SEQUENC	E LE	ENGTI	I: 97	4 ba	ase p	pairs	;					•				
40	SEQUENC	E TY	PE:	nucl	.eic	acid	ì										
	STRANDE	DNES	SS: C	loubl	.е												
	TOPOLOG	Y: 1	inea	ır													

50

45

ANTI-SENSE: No

ORGANISM: Hepatitis C virus

IMMEDIATE EXPERIMENTAL SOURCE CLONE: 015-2

5	GCGC	SATC	CT C	CA C	CT C	CA TO	CG T	GG G	AC C	AA A	TG T	GG A	AG T	GT C	rc a	TA C	GG 51
			D.	ro D	ro B	ro 6	~~ M·	en A	~~ <i>C</i>	1 . v.	- t . m.	T.	a	T	- .		
			F.	1 1	ro P	10 5	SL 1.	ւր <i>ռ</i> 5	вр С.	TII M	et T		ув С 10	Ås re	∋u 1.	Te A	rg
10	СТА	AAA	ССТ	_	CTA	CAC	GGG	_	ACA	ccc	ርጥር			ACC	www.y	CCA	99
					Leu												33
	15	•				20					25		-1-	9		30	
	GCC	GTT	CAA	AAC	GAG	GTC	ACC	CTC	ACA	CAC		ATA	ACC	AAG	TTC		147
15					Glu												
					35					40				-	45		
	ATG	GCA	TGC	ATG	TCG	GCT	GAC	CTA	GAG	GTC	GTC	ACT	AGC	ACC	TGG	GTG	195
	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	
20				50					55					60			
	CTG	GTA	GGC	GGG	GTC	CTC	GCA	GCT	CTG	GCC	GCG	TAC	TGC	CTG	ACA	ACG	243
	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	
			65					70	•				75				
25					ATC												291
	Gly		Val	Val	Ile	Val		Arg	Ile	Ile	Leu		Gly	Arg	Pro	Ala	
		80					85					90					
					AGG												339
30	95	TTE	PIO	Asp	Arg	100	vaı	ren	ıyr	Arg		Phe	Asp	Glu	Met		
		ጥርር	GCC	ጥሮል	CAC		ccc	ጥአሮ	አመሮ	CAA	105	CCA	a mc	CAC	CITIC	110	207
					His												301
		0,10			115			-1-	~~~	120	GIII	GLy	Me c	GIII	125	Ala	
35	GAG	CAA	TTC	AAG	CAG	AAG	GCG	CTC	GGT		TTG	CAA	ACA	GCT	_	AAG	435
					Gln												
				130		_			135					140			
	CAA	GCG	GAG	GCT	GCT	GCT	ccc	GTG	GTG	GAG	TCC	AAA	TGG	CGA	GCC	CTT	483
40					Ala												
			145					150					155	•			
	GAG	ACC	TTC	TGG	GCA	AAG	CAC	ATG	TGG	AAT	TTC	ATC	AGC	GGG	ATA	CAG	531
	Glu	Thr	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly	Ile	Gln	
45		160					165					170					

55

	TAC	TTG	GCA	GGC	TTG	TCC	ACT	CTG	CCT	GGA	AAC	ccc	GCG	ATA	GCA	TCA	579
	Tyr	Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	Ala	Ile	Ala	Ser	
5	175					180					185					190	
5	CTG	ATG	GCA	TTC	ACA	GCC	TCT	ATC	ACC	AGC	CCT	CTC	ACC	ACC	CAA	CAT	627
	Leu	Met	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	Thr	Thr	Gln	His	
	•				195					200					205		
															CTC		675
10	Thr	Leu	Leu	Phe	Asn	Ile	Phe	Gly	Gly	Trp	Val	Ala	Ala	Gln	Leu	Ala	
				210					215					220			
															GGC		723
45	Pro	Pro		Ala	Ala	Ser	Ala		Val	Gly	Ala	Gly	Ile	Ala	Gly	Ala	
15			225					230					235				
															CTG		771
	Ala		GIY	Ser	He	Gly		Gly	Lys	Val	Leu		Asp	Ile	Leu	Ala	
20	CCT	240	CCA	003	000	C/D/C	245	000			~~~	250					
20															GTC		819
	255	TYL	GIY	ATG	GIY	260	Ald	GTÅ	ATG	ren	265	Ala	Pne	тйв	Val		
		GGC	GAG	ልጥር	CCC		»CC	CAC	GNC	മസ്		220	mmx	OMO.	CCT	270	067
25															Pro		867
23		1			275	UUL	****	014	Au p	280	V 41	AOH	Бец	nea	285	ATG	
	ATC	CTC	TCT	ССТ		GCC	CTG	GTC	GTC		GTC	GTG	ጥርር	GC A	GCA	מיים	915
															Ala		713
30				290	•				295	1			-1-	300			
	CTG	CGT	CGA	CAT	GTG	GGC	CCA	GGG	GAG	GGG	GCT	GTG	CAG		ATG	AAC	963
															Met		
			305					310		_			315	-			
35	CGG	CTG	C AC	CC													974
	Arg	Leu															
		320															
40	SEQ ID																
	SEQUENC	E LE	ENGTH	f: 19	bas	e pa	irs										
	SEQUENC				eic	acid	l										
	TOPOLOG																
4 5	MOLECUI				r												
	ORIGINA	L SC	URCE	•													

ORGANISM: Hepatitis C virus CTCCACCATAGATCACTCC 19 5 SEQ ID NO: 106 SEQUENCE LENGTH: 18 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 10 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 15 AGGTCTAGTAGACCGTGC 18 SEQ ID NO: 107 SEQUENCE LENGTH: 18 base pairs 20 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA ORIGINAL SOURCE 25 ORGANISM: Hepatitis C virus AGGAAGACTTCCGAGCGG 18 30 SEQ ID NO: 108 SEQUENCE LENGTH: 19 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 35 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 40 CGTGAACTATGCAACAGGG 19 SEQ ID NO: 109 SEQUENCE LENGTH: 18 base pairs 45 SEQUENCE TYPE: nucleic acid

50

TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 10 ACCGCTCGGAAGTCTTCC 18 SEQ ID NO: 110 SEQUENCE LENGTH: 18 base pairs 15 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA 20 ORIGINAL SOURCE ORGANISM: Hepatitis C virus 25 GGGCAAGTTCCCTGTTGC 18 SEQ ID NO: 111 SEQUENCE LENGTH: 18 base pairs 30 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: CDNA 35 ORIGINAL SOURCE ORGANISM: Hepatitis C virus 40 GCTGGATTCTCTGAGACG 18 SEQ ID NO: 112 SEQUENCE LENGTH: 23 base pairs 45 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: CDNA 50 ORIGINAL SOURCE ORGANISM: Hepatitis C virus 55 GAGGCCGTGAACTGCGATGA 23

	SEQ ID NO: 113
	SEQUENCE LENGTH: 23 base pairs
5	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
10	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
	TTCTCTAAGGTGGCNTCNGCNTG 23
15	N: inosine
	GRO TR NO 114
	SEQ ID NO: 114
20	SEQUENCE LENGTH: 21 base pairs
	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
25	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
-30	CCGGACGCGTTGAANCTNGNGT 21
	N: inosine
	SEQ ID NO: 115
35	SEQUENCE LENGTH: 23 base pairs
	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
40	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
4 5	
	CATCCAGGTACAACCGAACCA 23
50	SEQ ID NO: 116
	SEQ ID NO: 116 SEQUENCE LENGTH: 24 base pairs
	SEQ ID NO: 116 SEQUENCE LENGTH: 24 base pairs SEQUENCE TYPE: nucleic acid
	SEQ ID NO: 116 SEQUENCE LENGTH: 24 base pairs

ORIGINAL SOURCE ORGANISM: Hepatitis C virus AACACACGGCCGCCNCANGGNAA N: inosine 10 SEQ ID NO: 117 SEQUENCE LENGTH: 19 base pairs SEQUENCE TYPE: nucleic acid 15 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 20 ORGANISM: Hepatitis C virus CCGGATCCCACAAGCCGTNGTNGA 25 N: inosine SEQ ID NO: 118 SEQUENCE LENGTH: 20 base pairs 30 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: CDNA 35 ORIGINAL SOURCE ORGANISM: Hepatitis C virus 40 GACATGCATGTCATGATGTA 20 SEQ ID NO: 119 SEQUENCE LENGTH: 26 base pairs 45 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA 50 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GGCTGCAGCCGGTTCATCCACTGCAC

SEQ ID NO: 120 SEQUENCE LENGTH: 26 base pairs SEQUENCE TYPE: nucleic acid 5 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 10 ORGANISM: Hepatitis C virus GCGGATCCTGCTTCGCCCAGAAGGTC 26 15 SEQ ID NO: 121 SEQUENCE LENGTH: 22 base pairs SEQUENCE TYPE: nucleic acid 20 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 25 ORGANISM: Hepatitis C virus GACACATGTGTTGCAGTCGATC 22 30 SEQ ID NO: 122 SEQUENCE LENGTH: 24 base pairs SEQUENCE TYPE: nucleic acid 35 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 40 ORGANISM: Hepatitis C virus CGGTCCNAGNAGTATCTCNTTNCC 24 N: inosine 45 SEQ ID NO: 123 SEQUENCE LENGTH: 35 base pairs 50 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: CDNA 55 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

ATGGGCCCGGGNGANAGNAGNCTCCCCCTNCTNT	C 35
N: inosine	

SEQ ID NO: 124

5

10

15

25

30

35

40

45

50

SEQUENCE LENGTH: 20 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

20 GGCTATACCGGCGACTTCGA 20

SEQ ID NO: 125

SEQUENCE LENGTH: 27 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGGATCCGGCCTCACCCACATAGATG 27

SEQ ID NO: 126

SEQUENCE LENGTH: 23 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGGATCCTCCACCTCCATCGTG 23

SEQ ID NO: 127

55 SEQUENCE LENGTH: 20 base pairs

5	SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus
10	CTGCTGTCGCCCNGNCCCAT 20 N: inosine
15	SEQ ID NO: 128 SEQUENCE LENGTH: 23 base pairs
20	SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA ORIGINAL SOURCE
25	ORGANISM: Hepatitis C virus
30	ATCACGTGGGGNGCAGANACNGC 23 N: inosine
	SEQ ID NO: 129 SEQUENCE LENGTH: 21 base pairs
35	SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA
40	ORIGINAL SOURCE ORGANISM: Hepatitis C virus
45	TGTGCCTGNTTNTGGATGATG 21 N: inosine
50	SEQ ID NO: 130 SEQUENCE LENGTH: 21 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear
55	MOLECULE TYPE: CDNA ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GGTGAGCATGGAGGTGACCAC 21

SEQ ID NO: 131

SEQUENCE LENGTH: 21 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

TCATCCTCCTCCGCTCGAAGC 21

SEQ ID NO: 132

SEQUENCE LENGTH: 23 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

30 ORIGINAL SOURCE

5

10

15

20

25

35

40

45

50

55

ORGANISM: Hepatitis C virus

GTGGACGCCTTNGCCTTCATNTC 23

N: inosine

SEQ ID NO: 133

SEQUENCE LENGTH: 21 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

ACGGATGTCNTTCTCNGTNAC 21

N: inosine

SEQ ID NO: 134

SEQUENCE LENGTH: 30 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 10 GGCGGAATTCCTGGTCATAGCCTCCGTGAA SEQ ID NO: 135 15 SEQUENCE LENGTH: 21 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 20 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 25 GGGGNATGGCCTATTGGCCTG N: inosine 30 SEQ ID NO: 136 SEQUENCE LENGTH: 21 base pairs SEQUENCE TYPE: nucleic acid 35 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 40 ORGANISM: Hepatitis C virus GGCATGTGGGCCCAGGGGAGG 21 45 SEQ ID NO: 137 SEQUENCE LENGTH: 20 base pairs SEQUENCE TYPE: nucleic acid 50 TOPOLOGY: linear MOLECULE TYPE: cDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus

TGTGAGCCCGAACCGGATGT 20

•	SEQ ID NO: 138
5	SEQUENCE LENGTH: 23 base pairs
	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
10	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
15	
	GTGGTANTCCTGGACTCNTTNGA 23
	N: inosine
20	
	SEQ ID NO: 139
	SEQUENCE LENGTH: 22 base pairs
	SEQUENCE TYPE: nucleic acid
25	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
30	ORGANISM: Hepatitis C virus
	ACTACCGNGACGTGCTNAANGA 22
35	N: inosine
	SEQ ID NO: 140
40	SEQUENCE LENGTH: 30 base pairs SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
4-	ORIGINAL SOURCE
45	ORGANISM: Hepatitis C virus
	oxdinion. Reputities C VIIIas
	TGGGGATCCCGTATGATACCCGCTGCTTTG 30
50	
	SEQ ID NO: 141
	SEQUENCE LENGTH: 24 base pairs
55	SEQUENCE TYPE: nucleic acid

	TOPOLOGY: linear	
E	MOLECULE TYPE: CDNA	
5	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
10	ATTGTCAGATCTACGGGGCCACTT 24	
	SEQ ID NO: 142	
15	SEQUENCE LENGTH: 43 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: CDNA	
20	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
25	GCAAGCTTAAAAAAAAAAAGGGGGATGGCCTATTGGCCTGGA	43
	SEQ ID NO: 143	
30	SEQUENCE LENGTH: 17 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
35	MOLECULE TYPE: CDNA	
33	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
40	GTAAAACGACGGCCAGT 17	
	SEQ ID NO: 144	
45	SEQUENCE LENGTH: 17 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA	
50	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
55	CAGGAAACAGCTATGAC 17	

	SEQ ID NO: 145
	SEQUENCE LENGTH: 35 base pairs
5	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
10	ORGANISM: Hepatitis C virus
	GCAAGCTTATGAGCACAAATCCAAAAACCCCAAAGA 35
15	
	SEQ ID NO: 146
	SEQUENCE LENGTH: 38 base pairs
20	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
25	ORGANISM: Hepatitis C virus
	GCGAATTCAGATCTTCACCTACGCCGGGGGTCCGTGGG 38
30	
	SEQ ID NO: 147
	SEQUENCE LENGTH: 39 base pairs
35	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
40	ORGANISM: Hepatitis C virus
	GCGAATTCAGATCTTCAGATTCTCTGAGACGGCCCTCGT 39
4 5	
	SEQ ID NO: 148
	SEQUENCE LENGTH: 17 base pairs
50	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
55	ORGANISM: Hepatitis C virus

GCTACTCCGGATACCAC 17

5	SEQ ID NO: 149	
v	SEQUENCE LENGTH: 35 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
10	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
15		
	GCGTCGACGCTAGCATGAGCACAAATCCAAAACCC 35	
	CRO ID NO. 150	
20	SEQ ID NO: 150	
	SEQUENCE LENGTH: 35 base pairs	
	SEQUENCE TYPE: nucleic acid	
25	TOPOLOGY: linear MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
30	ONGARISM: Nepatitis C Viius	
	GCGTCGACGCTAGCAGGTCTCGTAGACCGTGCATC 35	
35	SEQ ID NO: 151	
	SEQUENCE LENGTH: 40 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
40	MOLECULE TYPE: CDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
45		
	GCGAATTCGCTAGCTCAGGATTCTCTGAGACGGCCCTCGA	40
50	SEQ ID NO: 152	
JU	SEQUENCE LENGTH: 35 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
55	MOLECULE TYPE: CDNA	

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCAAGCTTATGCGGATCCCACAAGCCGTGGTGGAT 35

SEQ ID NO: 153

5

15

30

35

SEQUENCE LENGTH: 24 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

CGGATCCCACAAGCCGTGGTGGAT 24

SEQ ID NO: 154

²⁵ SEQUENCE LENGTH: 43 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGAATTCAGATCTTCATCACTCTAAGGTGGCGTCGGCGTGGG 43

SEQ ID NO: 155

40 SEQUENCE LENGTH: 11 base pairs

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

45 TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCAAGCTTATG 11

⁵⁵ SEQ ID NO: 156

	SEQUENCE LENGTH: 20 base pairs	
	SEQUENCE TYPE: nucleic acid	
5	STRANDEDNESS: double	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
10	ORGANISM: Hepatitis C virus	
	TGATGAAGATCTGAATTCGC 20	
15		
	SEQ ID NO: 157	
	SEQUENCE LENGTH: 34 base pairs	
20	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
25	ORGANISM: Hepatitis C virus	
	GCAAGCTTATGTTCAACGCGTCCGGATGTCCGGA 34	
30		
	SEQ ID NO: 158	
	SEQUENCE LENGTH: 23 base pairs	
35	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
40	ORGANISM: Hepatitis C virus	
	TTCAACGCGTCCGGATGTCCGGA 23	
45		
	SEQ ID NO: 159	
	SEQUENCE LENGTH: 43 base pairs	
50	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
55	ORGANISM: Hepatitic C virus	

GCGAATTCAGATCTTCATCAACAACCGAACCAGTTGCCCTGCG 43

	SEQ ID NO: 160	
5	SEQUENCE LENGTH: 34 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
10	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
15	-	
	GCAAGCTTATGATCGGGGGGGTCGGCAACAATAC 34	
	SEQ ID NO: 161	
20	SEQUENCE LENGTH: 23 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
25	MOLECULE TYPE: CDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
30		
	ATCGGGGGGTCGGCAACAATAC 23	
0.5	SEQ ID NO: 162	
35	SEQUENCE LENGTH: 43 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
40	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
45		
	GCGAATTCAGATCTTCATCAAAGCTCTGATCTATCCCTGTCCT	43
50	SEQ ID NO: 163	
-	SEQUENCE LENGTH: 41 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	

MOLECULE TYPE: cDNA

ORIGINAL SOURCE

	ORGANISM: Hepatitis C virus	
5	GCGTCGACGCTAGCATGCGGATCCCACAAGCCGTGGTGGAT	41
	SEQ ID NO: 164	
10	SEQUENCE LENGTH: 40 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
15	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
20	GCGTCGACGCTAGCATGTTCAACGCGTCCGGATGTCCGGA	40
	SEQ ID NO: 165	
25	SEQUENCE LENGTH: 40 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
30	MOLECULE TYPE: CDNA	
*	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
35	GCGTCGACGCTAGCATGATCGGGGGGGTCGGCAACAATAC	40
	SEQ ID NO: 166	
40	SEQUENCE LENGTH: 40 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
45	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
50	GCGAATTCGCTAGCTCACTCTAAGGTGGCGTCGGCGTGGG	40
	SEQ ID NO: 167	
55	SEQUENCE LENGTH: 40 base pairs	

	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
5	MOLECULE TYPE: CDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
10	GCGAATTCGCTAGCTCAACAACCGAACCAGTTGCCCTGCG	40
	SEQ ID NO: 168	
15	SEQUENCE LENGTH: 40 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
20	MOLECULE TYPE: CDNA	
20	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
25	GCGAATTCGCTAGCTCAAAGCTCTGATCTATCCCTGTCCT	40
	SEQ ID NO: 169	
30	SEQUENCE LENGTH: 32 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
35	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
40	GCAAGCTTATGTGGTTGTGGATGATGCTGCTG 32	
	SEQ ID NO: :170	
45	SEQUENCE LENGTH: 21 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
50	MOLECULE TYPE: CDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
55	TGGTTGTGGATGATGCTGCTG 21	

	SEQ ID NO: 171
	SEQUENCE LENGTH: 44 base pairs
5	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: cDNA
	ORIGINAL SOURCE
10	ORGANISM: Hepatitis C virus
	GCGAATTCAGATCTTCATCACCTCCGGGCGGAGACNGGNAGNCC 44
15	N: inosine
	SEQ ID NO: 172
20	SEQUENCE LENGTH: 31 base pairs
	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
25	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
30	GCAAGCTTATGGGCAACGAGNTNCTNCTNGG 31
	N: inosine
35	SEQ ID NO: 173
	SEQUENCE LENGTH: 20 base pairs
	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
40	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
	ORGANISM: Hepatitis C virus
45	
	GGCAACGAGNTNCTNCTNGG 20
	N: inosine
50	
	SEQ ID NO: 174
	SEQUENCE LENGTH: 41 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus GCGAATTCAGATCTTCATCACTTCAGCCGTATGAGACACTT 41 10 SEQ ID NO: 175 SEQUENCE LENGTH: 31 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 15 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 20 GCAAGCTTATGCTGTCGCCCGGGCCCATCTC 31 25 SEQ ID NO: 176 SEQUENCE LENGTH: 20 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 30 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 35 CTGTCGCCCGGGCCCATCTC 40 SEQ ID NO: 177 SEQUENCE LENGTH: 41 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 45 MOLECULE TYPE: cDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 50 GCGAATTCAGATCTTCATCAACATGTGTTGCAGTCGATCAC

SEQ ID NO: 178

SEQUENCE LENGTH: 32 base pairs SEQUENCE TYPE: nucleic acid TOPOLOGY: linear 5 MOLECULE TYPE: CDNA ORIGINAL SOURCE ORGANISM: Hepatitis C virus 10 GCAAGCTTATGGGCTATACCGGNGACTTNGAC 32 N: inosine 15 **SEQ ID NO: 179** SEQUENCE LENGTH: 21 base pairs SEQUENCE TYPE: nucleic acid 20 TOPOLOGY: linear MOLECULE TYPE: CDNA ORIGINAL SOURCE 25 ORGANISM: Hepatitis C virus GGCTATACCGGNGACTTNGAC 21 N: inosine 30 SEQ ID NO: 180 SEQUENCE LENGTH: 35 base pairs 35 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA 40 ORIGINAL SOURCE ORGANISM: Hepatitis C virus GCGAATTCAGATCTTCAGTGCTTCGCCCAGAAGGT 45 35 SEQ ID NO: 181 SEQUENCE LENGTH: 29 base pairs 50 SEQUENCE TYPE: nucleic acid TOPOLOGY: linear MOLECULE TYPE: cDNA 55 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGCTAGCATGTGGTTGTGGATGATGCTG 29

SEQ ID NO: 182

SEQUENCE LENGTH: 38 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

15 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGAATTCGCTAGCTCACAGCCGGTTCATCCACTGCAC 38

SEQ ID NO: 183

SEQUENCE LENGTH: 32 base pairs

25 SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

30 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCAAGCTTATGCAGCGTGGGTACAAGGGGGTT 32

SEQ ID NO: 184

SEQUENCE LENGTH: 47 base pairs

40 SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

45 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGAATTCAGATCTTCATCAGAGCTGTGACCCAACCGTATATTGGTT 47

SEQ ID NO: 185

SEQUENCE LENGTH: 33 base pairs

55 SEQUENCE TYPE: nucleic acid

	TOPOLOGY: linear	
5	MOLECULE TYPE: CDNA	
•	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
10	GCGCTAGCATGGGGTACAAGGGGGTTTGGCGGG	33
	SEQ ID NO: 186	
15	SEQUENCE LENGTH: 32 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
20	MOLECULE TYPE: cDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
25	GCGCTAGCTCATCGGTTGGGGAGCAGGTAGAT	32
	SEQ ID NO: 187	
30	SEQUENCE LENGTH: 26 base pairs	
•	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
35	MOLECULE TYPE: cDNA	
••	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
40	GGATCCCCCAAGCTTGGGGGAATTC 26	,
•	SEQ ID NO:188	
45	SEQUENCE LENGTH: 31 base pairs	
	SEQUENCE TYPE: nucleic acid	
	TOPOLOGY: linear	
50	MOLECULE TYPE: CDNA	
	ORIGINAL SOURCE	
	ORGANISM: Hepatitis C virus	
55	AGCTTACTAGTTAATACGACTCACTATAGGG	31

	SEQ ID NO:189
	SEQUENCE LENGTH: 33 base pairs
5	SEQUENCE TYPE: nucleic acid
5	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
10	ORGANISM: Hepatitis C virus
	CTGGCACCCTATAGTGAGTCGTATTAACTAGTA 33
15	
	SEQ ID NO:190
	SEQUENCE LENGTH: 44 base pairs
20	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
25	ORGANISM: Hepatitis C virus
	TGCCAGCCCCTGATGGGGGCGACACTCCACCATAGATCACTCC 44
30	
	SEQ ID NO:191
	SEQUENCE LENGTH: 45 base pairs
35	SEQUENCE TYPE: nucleic acid
	TOPOLOGY: linear
	MOLECULE TYPE: CDNA
	ORIGINAL SOURCE
40	ORGANISM: Hepatitis C virus
	TCACAGGGGAGTGATCTATGGTGGAGTGTCGCCCCCATCAGGGGG 45
45	
	SEQ ID NO:192
	SEQUENCE LENGTH: 40 base pairs
50	SEQUENCE TYPE: nucleic acid
JU	TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORGANISM: Hepatitis C virus

ORIGINAL SOURCE

CCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAGC 40

SEQ ID NO:193

SEQUENCE LENGTH: 37 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE

5

15

20

25

35

40

45

50

ORGANISM: Hepatitis C virus

CATGGCTAGACGCTTTCTGCGTGAAGACAGTAGTTCC 37

SEQ ID NO:194

SEQUENCE LENGTH: 33 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

30 ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCAAGCTTATGCTGCTGTCGCCCGGGCCCATCT 33

SEQ ID NO:195

SEQUENCE LENGTH: 38 base pairs

SEQUENCE TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: CDNA

ORIGINAL SOURCE

ORGANISM: Hepatitis C virus

GCGAATTCAGATCTTCATCATGTGTTGCAGTCGATCAC 38

Claims

- 55 1. An isolated gene encoding a polypeptide originated from h patitis C virus, wherein said polypeptide has an amino acid sequence of SEQ ID NO 101.
 - 2. An isolated gene encoding a polypeptide originat d from hepatitis C virus, wherein said polypeptide

has an amino acid sequence of SEQ ID NO 102.

- An isolated DNA originated from hepatitis C virus, wherein said DNA has a base sequence of SEQ ID NO 101.
- An isolated DNA originated from hepatitis C virus, wherein said DNA has a base sequence of SEQ ID NO 102.
- A polypeptide which comprises 115 amino acids from No. 1 to No. 115 of amino acid sequence of SEQ
 ID NO 3 or 7.
 - 6. An isolated DNA which encodes a polypeptide of Claim 127.
- 7- A polypeptide of 10 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 6 amino acids from No. 182 to No. 187 of amino acid sequence of SEQ ID NO 31 or 32.
 - 8. A polypeptide of 10 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 8 amino acids from Nos. 202 to 209 of amino acid sequence of SEQ ID NO 31 or 32.
 - A polypeptide which comprises 106 amino acids from No. 109 to No. 214 of amino acid sequence of SEQ ID NO 31 or 32.
- 25 10. A polypeptide which comprises 92 amino acids from No. 233 to No. 324 of amino acid sequence of SEQ ID NO 31 or 32.
 - 11. A polypeptide of 10 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 5 amino acids from No. 252 to No. 256 of amino acid sequence of SEQ ID NO 31 or 32.
 - 12. A polypeptide of 10 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 7 amino acids from No. 273 to No. 279 of amino acid sequence of SEQ ID NO 31 or 32.
 - 13. A polypeptide of 10 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 7 amino acids from No. 136 to No. 142 of amino acid sequence of SEQ ID NO 31 or 32.
- 40 14. A polypeptide of 17 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 31 or 32, wherein said polypeptide comprises at least 17 amino acids from No. 53 to No. 69 of amino acid sequence of SEQ ID NO 31 or 32.
- 15. A polypeptide which comprises all or 266 amino acids from No. 461 to No. 726 of amino acid sequence of SEQ ID NO 43.
 - 16. A polypeptide which comprises all or 42 amino acids from No. 963 to No. 1004 of amino acid sequence of SEQ ID NO 43.
- 50 17. A polypeptide which comprises all or 45 amino acids from No. 283 to No. 327 of amino acid sequence of SEQ ID NO 43.
 - 18. A polypeptide which comprises all or 74 amino acids from No. 477 to No. 550 of amino acid sequence of SEQ ID NO 43.
 - 19. A polypeptide which comprises 61 amino acids from No. 215 to No. 275 of amino acid sequence of SEQ ID NO 43.

55

20

30

- A polypeptide which comprises all or 74 amino acids from No. 413 to No. 486 of amino acid sequence of SEQ ID NO 75.
- 21. A polypeptide which comprises all or 997 amino acids from No. 415 to No. 1411 of amino acid sequence of SEQ ID NO 75.
 - 22. A polypeptide which comprises all or 19 amino acids from No. 247 to No. 265 of amino acid sequence of SEQ ID NO 75.
- 23. A polypeptide which comprises all or 74 amino acids from No. 655 to No. 728 of amino acid sequence of SEQ ID NO 75.
 - 24. A polypeptide which comprises all or 54 amino acids from No. 763 to No. 816 of amino acid sequence of SEQ ID NO 75.
 - 25. A polypeptide shown by at least 20 amino acid residues from No. 324 to No. 343 of amino acid sequence of SEQ ID NO 75, wherein said polypeptide comprises at least 8 amino acids.
- 26. A polypeptide which comprises all or 98 amino acids from No. 858 to No. 955 of amino acid sequence of SEQ ID NO 75.
 - 27. A polypeptide of 14 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 75, wherein said polypeptide comprises at least 14 amino acids from No. 356 to No. 369 of amino acid sequence of SEQ ID NO 75.
 - 28. A polypeptide which comprises all or 92 amino acids from No. 1009 to No. 1100 of amino acid sequence of SEQ ID NO 75.
- 29. A polypeptide which comprises all or 66 amino acids from No. 1160 to No. 1225 of amino acid sequence of SEQ ID NO 75.
 - **30.** A polypeptide of 18 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 75, wherein said polypeptide comprises at least 18 amino acids from No. 584 to No. 601 of amino acid sequence of SEQ ID NO 75.
 - 31. A polypeptide which comprises 42 amino acids from No. 615 to No. 656 of amino acid sequence of SEQ ID NO 75.
- 32. A polypeptide of 11 to 25 amino acid residues derived from amino acid sequence of SEQ ID NO 75, wherein said polypeptide comprises at least 11 amino acids from No. 326 to No. 337 of amino acid sequence of SEQ ID NO 75.
 - 33. A single-stranded DNA fragment or an antisense DNA fragment thereof which contains at least 15 nucleotides selected from 317 nucleotides from No. 1 to No. 317 of SEQ ID NO 1, 9, 11 or 12.
 - 34. The DNA fragment of Claim 221 comprising 16 to 30 base pairs.
 - 35. The DNA fragment of Claim 221 comprising 17 to 23 base pairs.
- 50 36. The use of a DNA and/or a polypeptide as claimed in any of the preceding claims for the preparation of a vaccine against hepatitis C virus.
 - 37. The use of a DNA and/or a polypeptide as claimed in many of the preceding claims for the serodiagnosis of hepatitis C related diseases.
 - 38. The use of a DNA as claimed in any of the preceding claims for a <u>in vitro</u> and/or <u>in vivo</u> screening system for a substance capable of specifically suppressing or controlling a proteolytic processing of a precursor protein of hepatitis C virus.

305

55

45

15